

Algorithmischer Journalismus

Eine Analyse der automatisierten Textproduktion im Journalismus
auf gesellschaftlicher, organisatorischer und professioneller Ebene

Abhandlung (kumulative Dissertation)
zur Erlangung der Doktorwürde
der Philosophischen Fakultät
der
Universität Zürich

vorgelegt von
Konstantin Nicholas Dörr

Angenommen im Frühjahrssemester 2017
auf Antrag der Promotionskommission:

Prof. Dr. Michael Latzer (hauptverantwortliche Betreuungsperson)
Prof. Dr. Ralf Hohlfeld

Zürich, 2017

Zusammenfassung

Die Dissertation verortet den Einfluss der automatisierten Textproduktion in Anlehnung an das Kontext-Modell von Weischenberg auf gesellschaftlicher, organisatorischer und professioneller Ebene des Journalismus. Die Analyse ist eingebettet in den sozio-technischen Kontext der Infrastruktur und Entwicklung des Internets, in den Kontext von Internetanwendungen auf Basis algorithmischer Selektion, der Rolle von Technik als Akteur und Institution sowie in den Kontext von Big Data. Neben einer Forschungsübersicht und einer Systematisierung der Technik im Forschungsfeld des Computational Journalism bilden eine Technikpotentialanalyse, Interviews mit Software-Anbietern und Journalisten sowie zwei theoretische Beiträge zu normativen und ethischen Herausforderungen die Eckpfeiler dieser Arbeit. Die Potentialanalyse und das Konzept von Technik als Akteur erlauben das Wechselspiel von Mensch und Software zu operationalisieren und der Technik Leistungen des Informationsjournalismus zuzuschreiben. Dadurch werden die spezifischen Handlungsebenen auch für normative und ethische Anschlussfragen zugänglich. Diese werden aus der Perspektive der journalistischen Ethik diskutiert. Mit der Befragung von Software-Anbietern können zudem erste Rückschlüsse auf den Markt gezogen werden. Dabei zeigen die Ergebnisse, dass er sich in einer frühen Expansionsphase befindet. Die identifizierten Herausforderungen werden zudem aus praktischer Perspektive durch eine Journalistenbefragung kontextualisiert.

Abstract

This thesis investigates the impact of automated text production at the social, organizational and professional levels of journalism. The context model of Weischenberg serves as a heuristic framework to analyze the technical, economic and ethical challenges associated with this technology. The analysis is embedded in the socio-technical context of the infrastructure and development of the Internet, of internet applications based on algorithmic selection, of the role of technology as an actor and institution, and the context of big data. In addition to a literature review and a systematization of automated text production in computational journalism, the cornerstones of this work are a technical potential analysis, expert interviews with software providers and journalists, as well as two theoretical contributions on normative and ethical challenges. The thesis helps to operationalize the interplay between human actors and software, showing how the technology can fulfill functions of journalism after initial programming. This approach also allows the visualization and discussion of normative and ethical questions, which were discussed in relation to journalism ethics. The survey of software providers allows initial conclusions to be drawn about the market for automated text production. Results show that it is at an early phase of expansion. Potentials and limitations of the technology are also contextualized from a practical perspective by a survey of journalists.

Vorwort und Danksagung

In Anbetracht der Tatsache, dass das Vorwort bei einer Dissertation darüber entscheiden kann, ob der Leser auch zum restlichen Inhalt weiterblättert, kommt ihm eine nicht zu unterschätzende Bedeutung zu. Das Vorwort funktioniert wie der Teaser eines Buzz-Feed-Textes: «Willst du wissen, wie der Autor den langen und beschwerlichen Entstehungsprozess der Arbeit gemeistert hat ... 10 krasse Dinge, mit denen du nie gerechnet hättest.»

Wer allerdings denkt, dass nun gefühlsecht genau darüber philosophiert wird, wird – wie so oft bei einem Clickbait – enttäuscht. Stattdessen soll an dieser Stelle besonders jenen gedankt werden, die mich während dieser Lebensphase aufopfernd begleitet, kritisiert, aufgeheitert und motiviert haben.

Ich danke von Herzen Michael Latzer, Natascha Just, Ralf Hohlfeld, Julian Wallace, Moritz Büchi, Neil Thurman, Nina Köberer, Florian Saurwein, Alexander Godulla, meinen KoautorInnen und KollegInnen am IPMZ sowie insbesondere Johanna Martens.

Nicht zuletzt gilt mein größter Dank meinen Eltern, die mir immer vermittelt haben, dass Bildung das höchste Gut ist. Auch ein abgebrochenes Jura-Studium ließ Sie nie an diesem Grundsatz zweifeln.

Konstantin Nicholas Dörr

Zürich, im Mai 2017

I. Einleitung: Problemstellung, Ziele und Aufbau der Arbeit	1
II. Theoretische Zugänge	7
1. Verortung im Forschungsprogramm «Algorithmische Selektion im Internet»	8
2. Technik als Akteur und Institution	11
3. Das Kontext-Modell des Journalismus	16
3.1 Ebene der Mediensysteme	18
3.2 Ebene der Medieninstitutionen	21
3.3 Ebene der Medienaussagen	25
3.4 Ebene der Medienakteure	26
4. Zusammenfassung und Überblick	29
III. Methodische Zugänge	35
1. Die Fallstudie	36
1.1 Potentialanalyse	36
1.2 Experteninterview	37
1.3 Beobachtung	38
2. Zusammenfassung und Diskussion	39
IV. Diskussion der Forschungsergebnisse	41
1. Einfluss auf die Ebene der Mediensysteme	41
2. Einfluss auf die Ebene der Medieninstitutionen	45
3. Einfluss auf die Ebene der Medienaussagen	47
4. Einfluss auf die Ebene der Medienakteure	50
5. Zusammenfassung	54
V. Abschließende Betrachtung und Forschungsausblick	57
VI. Publikationen	61
Artikel I – Algorithmen, Big Data und ihre Rolle im Computational Journalism	61
Artikel II – Mapping the Field of Algorithmic Journalism	79
Artikel III – When Reporters get Hands-On with Robo-Writing	103
Artikel IV – Normative Qualitätsansprüche an algorithmischen Journalismus	125
Artikel V – Ethical Challenges of Algorithmic Journalism	139
VII. Abbildungs- und Tabellenverzeichnis	157
VIII. Literaturverzeichnis	159

Anhang

A. Interviews und Transkripte	181
A.1 Manfred Stede, Prof. für Angewandte Computerlinguistik, Universität Potsdam	181
A.2 Jens Wagner, Leiter Technik und Projekte, Sport-Informationen-Dienst (SID)	185
A.3 Carsten Erdmann, Chefredakteur, Berliner Morgenpost	190
A.4 Tom Kent, Deputy Managing/Standards Editor, The Associated Press	204
B. Ausweis der Eigenleistung	211
C. Lehr- und Forschungstätigkeiten	213
C.1 Lehre	213
C.2 Präsentationen, Panel und Workshops	214
C.3 Weitere Publikationen während des Doktorats	215
D. Curriculum Vitae	217

I. Einleitung: Problemstellung, Ziele und Aufbau der Arbeit

«Nimm nie einen Menschen, wenn du eine Maschine dafür nehmen kannst.»

Agent Smith, The Matrix (1999)

Das Schreckgespenst einer Übermacht der Maschinen geistert nicht nur durch zahlreiche Science-Fiction-Filme, sondern wird aktuell auch im Rahmen der fortschreitenden Digitalisierung und Automatisierung in verschiedenen Industriezweigen sichtbar (vgl. Acemoglu & Restrepo 2016; Arntz, Gregory, & Zierhan 2016; Autor 2015). Diese Prozesse setzen betroffene Akteure unter Druck. Sie müssen auf die Entwicklung reagieren und gegebenenfalls Strukturen und Handlungsorientierungen an die neuen technischen Möglichkeiten und Erfordernisse anpassen. Dies führt mitunter zu neuen Kooperations- und Konkurrenzbeziehungen sowie zu neuen Handlungsmustern und Verantwortlichkeiten im Verhältnis von Mensch und Technik (vgl. Dolata & Werle 2007: 35ff.). Auch die Geschichte des Journalismus ist eng mit der Entwicklung der Technik verknüpft (vgl. Stöber 2008, 2013). In den letzten Jahren steht vermehrt der «Algorithmus» im Scheinwerferlicht der kommunikationswissenschaftlichen Aufmerksamkeit (für einen Überblick über die praktischen Herausforderungen vgl. Rainie & Anderson 2017). Algorithmen klassifizieren, strukturieren und filtern – je nach Funktion der Anwendungen – digitale Informationen, weisen ihnen auf Basis programmierter Regeln Relevanz zu und entwickeln dabei eine Eigendynamik (vgl. Rieder 2004). Diese Prozesse verändern die Recherche, Auswahl, Darstellung, Distribution und Rezeption von Inhalten und damit auch die Art und Weise, wie Journalismus die Gesellschaft beobachtet und beschreibt (vgl. u.a. Flew et al. 2012; Napoli 2014; Neuberger 2008; Neuberger & Nuernbergk 2015). Die Leistungen des Computers bestehen damit nun nicht mehr aus «vollkommen determinierten Operationen» (Esposito 1993: 351).

Die bis jetzt letzte Evolutionsstufe im Zusammenspiel von Mensch und Technik wurde mit dem Einsatz der automatisierten Texterstellung im Journalismus erreicht. Die technische Grundlage basiert dabei auf dem Prinzip der natürlich-sprachlichen Inhaltserstellung, der Natural Language Generation (NLG), das in dieser Arbeit als algorithmischer Journalismus konzeptualisiert wird (zur Definition siehe insbesondere Artikel II).¹ Erste Versuche, Texte automatisch zu produzieren, fanden bereits in den 1960er Jahren statt. Damals wurde mit Ansätzen einer Repräsentation von Wissen mithilfe semantischer Netzwerke bzw. Netze operiert (Überblick siehe Carstensen 2012: 169ff.). Der «Statistik-Boom»

¹ Die Begriffe «automatisierte» bzw. «algorithmische» «Textgenerierung» oder «Inhaltserstellung» rufen nicht zwangsläufig nach einem journalistischen Verwendungszweck. In dieser Arbeit werden sie aber synonym mit dem Begriff «algorithmischer Journalismus» verwendet, da hier der Einfluss der Technik nur im Hinblick auf einen journalistischen Verwendungszweck untersucht wird.

(Stede, Anhang A1) im Forschungsfeld, die Fortschritte auf dem Gebiet künstlichen Intelligenz² (KI), die zunehmende Verfügbarkeit digitaler Daten sowie die finanzielle Schieflage vieler Medienorganisationen haben schließlich zu einem Einsatz im Journalismus geführt.

Die Analyse des Wechselspiels zwischen Mensch und Technik, zwischen Journalist und Algorithmus, ist besonders dann relevant, wenn sich Leistungen des Journalismus aufgrund veränderter Produktionsbedingungen wandeln. Dabei ist diese Entwicklung für eine Verortung kommunikationswissenschaftlicher Fragestellungen in Bezug auf die Felder und Funktionen gesellschaftlicher Kommunikation zentral (vgl. Latzer et al. 2014). Insbesondere für den Journalismus, dessen Kernleistung sich auf die Ermöglichung gesellschaftlicher Kommunikation bezieht, ist es notwendig, dieses Phänomen interdisziplinär und ansatzoffen zu untersuchen (vgl. Hanitzsch & Engesser 2014; Pöttker 2010; Quandt 2000). Es stellt sich somit einerseits die Frage, welche Leistungen algorithmischer Journalismus erbringen kann, und andererseits die Frage, wie man diese Leistungen und Herausforderungen operationalisierbar und analysierbar macht (vgl. Neuberger 2009: 26; Rammert 2003, 2007; zur KI vgl. Kurzweil 1990). Dazu nutzt diese Dissertation verschiedene theoretische und empirische Zugänge. Die Analyse dieser Prozesse ist eingebettet in den sozio-technischen Kontext der Infrastruktur und Entwicklung des Internets als Informations- und Kommunikationstechnologie (vgl. Sandvig 2013), in den Kontext von Internetanwendungen auf Basis algorithmischer Selektion (vgl. Latzer et al. 2014) und der Rolle von Technik als Akteur und Institution (Dolata & Werle 2007: 17ff.), das Kontext-Modell des Journalismus nach Weischenberg (1992) sowie den Kontext des Phänomens «Big Data», das in der Konsequenz zu einer «Verdatung der Gesellschaft» (Neuberger & Nuernbergk 2015: 200) führt.

Die Thesis betreibt dabei explorative Grundlagenforschung und hilft algorithmischen Journalismus im Forschungsbereich des Computational Journalism zu verorten und seinen Einfluss auf den Journalismus auf gesellschaftlicher, organisatorischer und professioneller Ebene zu analysieren. Um dies zu tun, greift die Arbeit auf das Kontext-Modell von Weischenberg (1992) zurück. Mit diesem Modell lassen sich die Einflussfaktoren im Bereich des journalistischen Handelns und der Einfluss von algorithmischem Journalismus auf der Ebene der Mediensysteme (gesellschaftliche Ebene; Normenkontext), der Medieninstitutionen (organisatorische Ebene; Strukturkontext), der Medienaussagen (Funktionskontext) und der Ebene der Medienakteure (professionelle Ebene; Rollenkontext) verorten und untersuchen (vgl. Weischenberg 1992: 68; Weischenberg et al. 2006a,b; siehe Abb. 1).³ Die systemtheoretische Logik dient hier in erster Linie zur Lokalisierung der empirischen und theoretischen Befunde der Dissertation (vgl. Esser 1998: 23). Die Ebenen sind dabei in

² Maschinelle oder auch künstliche Intelligenz wird von Minsky (1988; zitiert nach Whitby 1996: 20) definiert als «the science of making machines do things that would require intelligence if done by men».

³ Zur Kritik am Modell, u.a. Baum (1994: 361ff.), Neuberger (1996: 185ff.), Esser (1998).

Anlehnung an Schimank (1996: 243ff.) als «offene Orientierungshorizonte» zu verstehen. Sie stehen in einem engen Interaktionsverhältnis und beeinflussen sich gegenseitig. Die Frage nach der Funktion von algorithmischem Journalismus wird dabei im Kontext des Bezugssystems Journalismus diskutiert (vgl. Beck 2013: 95). Journalismus wird hier ausschließlich in seiner Funktion als Informationsjournalismus – nicht als Unterhaltungsjournalismus – operationalisiert und als Leistungssystem innerhalb des Funktionssystems Öffentlichkeit verstanden. Er beobachtet Teilbereiche der Gesellschaft und stellt der Öffentlichkeit gesellschaftlich relevante, nichtfiktionale Themen und Aussagen mit einer bestimmten Relevanz zur Verfügung, die nach professionellen und ethischen Standards innerhalb der gesellschaftlichen Rahmenordnung produziert werden (vgl. Neuberger 1996; Scholl & Weischenberg 1998: 63ff.). Journalismus liefert dabei nicht nur ein Abbild der Realität, sondern ist auch zentral an der Wirklichkeitsrekonstruktion und -konstitution beteiligt (vgl. Tuchman 1973: 129; zur algorithmischen Realitätskonstruktion Just & Latzer 2016).

Um die Wirkungsweise von algorithmischem Journalismus innerhalb des Journalismus theoretisch und empirisch greifbar zu machen, wird neben dem Kontext-Modell als kommunikationswissenschaftlichem Analyserahmen auch auf das techniksoziologische Konzept der Rolle von Technik als Akteur und Institution zurückgegriffen. Technische Innovationen wie der algorithmische Journalismus werden hier allgemein als Neuerungen konzeptualisiert, die den Kern des Medienwandels bilden, und gleichzeitig als Triebkräfte, die die soziale und ökonomische Dynamik von Gesellschaften verändern (vgl. Latzer 2009; 2013). Dieses koevolutionäre Verständnis und die Einbindung der techniksoziologischen Perspektive erlauben die Integration des technischen Wandels auf gleicher Ebene mit ökonomischen, politischen und kulturellen Triebkräften als «individuelle, organisationale und gesellschaftliche Veränderungsprozesse» (Dogruel 2013: 368; Latzer 2013; Young & Hermida 2015: 383). Diese Beziehung umfasst sowohl Akteursstrategien und Handlungen von Technik, Journalisten und Software-Anbietern als auch institutionelle und strukturelle Kontexte, innerhalb derer algorithmischer Journalismus entwickelt und genutzt wird (vgl. Dolata & Werle 2007: 16f.; Matsuzaki 2011; Neuberger 2009; Rip 2007). Insbesondere im Internet kann dieser Wandel des Journalismus beobachtet und untersucht werden (vgl. Meier 2011b).

Die Dissertation fokussiert dabei auf das Wechselspiel dieser Veränderungsprozesse und überwindet die Dichotomie von technik- und sozialdeterministischen Interpretationen des Einflusses von Technik bei der Beobachtung und Beschreibung der Gesellschaft im Journalismus (vgl. Dolata & Werle 2007: 13). Die Studie geht daher in ihren Beschreibungs- und Erklärungsversuchen von einem «Nebeneinander von Kommunikation und Handlung, von System und Akteur, von Autonomie und Fremdsteuerung» aus (Hohlfeld

2003: 107). Damit ist es möglich, die Rolle von Technik im Medienwandel und ihren Einfluss auf soziales Handeln, Strukturen und Institutionen des Journalismus differenziert zu betrachten (vgl. Lewis & Westlund 2015a,b; Werle 2005).

Methodisch greift die Dissertation dabei auf den Forschungsansatz der Fallstudie zurück. Einzelne Fallanalysen eignen sich besonders für die explorative empirische und theoretische Analyse bisher wenig erforschter Phänomene wie des algorithmischen Journalismus, der hier mit einem breiten sozialwissenschaftlichen Zugang und vor dem Hintergrund seiner Kontextbezogenheit auf der Makro-, Meso- und Mikroebene des Journalismus analysiert wird (vgl. Häder 2010: 353f.; Weischenberg 1992). Denn erst diese Verknüpfung von Makro-, Meso- und Mikroperspektive ermöglicht die Identifizierung und Analyse von handlungsprägenden und strukturbildenden Einflussfaktoren (vgl. Wyss & Keel 2010: 341). Wenngleich die theoretischen Zugänge in den einzelnen Artikeln separat diskutiert werden, ist dieses Verständnis Klammer und Startpunkt für die Synthese und Diskussion der Forschungsergebnisse der Artikel I bis V.

Vor dem Hintergrund dieser unterschiedlichen Bezugspunkte und Aggregationsebenen lautet die forschungsleitende Frage für die theoretischen und empirischen Untersuchungen in dieser Arbeit:

Welchen Einfluss hat algorithmischer Journalismus auf gesellschaftlicher, organisatorischer und professioneller Ebene des Journalismus und welche Interdependenzen lassen sich als Konsequenz zwischen diesen Ebenen identifizieren?

Die Synopse gliedert sich dabei in folgende Kapitel: In Kapitel II – dem theoretischen Teil der Arbeit – wird algorithmischer Journalismus in das Forschungsprogramm «Algorithmische Selektion im Internet» eingebettet und als strukturierende Innovation operationalisiert, die technische, ethische und ökonomische Herausforderungen hervorruft (vgl. Latzer et al. 2014). Als konzeptionelle Linse für die Analyse der Wirkungsweise von algorithmischem Journalismus wird die techniksoziologische Perspektive von Technik als Akteur und Institution herangezogen (Kapitel II, Abschnitt 2). Dieses Verständnis erlaubt die Integration und Verknüpfung der Rolle von Technik mit dem kommunikationswissenschaftlichen Analyserahmen des Kontext-Modells von Weischenberg (1992) (Kapitel II, Abschnitt 3). Dieses Modell ermöglicht vor dem Hintergrund seiner Kontextbezogenheit die Verknüpfung der Forschungsfragen der Artikel I bis V auf gesellschaftlicher, organisatorischer und professioneller Ebene (vgl. Weischenberg et al. 2006a,b; siehe Abb. 1). Im Rahmen dieses Modells werden auf gesellschaftlicher Ebene insbesondere die ethischen und normativen Fragestellungen und Zugänge der Artikel IV und V vorgestellt. Auf organisatorischer Ebene werden technische, organisatorische und ökonomische Imperative zum Einfluss der Technik diskutiert (vgl. Artikel II und III). In der Anbieterbefragung in Artikel

II wird sowohl das ökonomische Potential von algorithmischem Journalismus mit der Einordnung in Marktphasen nach Heuss (1965) evaluiert, als auch das technische Leistungsspektrum mit der Potentialanalyse. Die Ergebnisse und Diskussion der ethischen, ökonomischen, technischen und organisatorischen Herausforderungen werden zudem auf professioneller Ebene mit der Journalistenbefragung gespiegelt (vgl. Artikel III). Das Kapitel schließt mit einer Übersicht über die zentralen Fragestellungen und den wissenschaftlichen Beitrag der einzelnen Artikel (vgl. Tabelle 1).

In Kapitel III wird die Fallanalyse als methodischer Zugang der Arbeit vorgestellt. Dieser differenziert sich in Technikpotentialanalyse (Artikel II), Experteninterviews (Artikel II und III) und Beobachtung (Artikel III). Eine Auseinandersetzung mit diesen methodischen Zugängen und den gewonnenen Implikationen für die Diskussion der Forschungsergebnisse runden das Kapitel ab.

Mit Verweis auf die übergeordnete Forschungsfrage folgt die Kontextualisierung der Forschungsergebnisse im Rahmen des Modells von Weischenberg (1992) auf gesellschaftlicher, organisatorischer und professioneller Ebene sowie unter Berücksichtigung der techniksoziologischen Perspektive in Bezug auf die Rolle von Technik (Kapitel IV). Die Arbeit mündet schließlich in einer Zusammenfassung und einem Ausblick für die zukünftige Forschung (Kapitel V).

II. Theoretische Zugänge

In diesem Kapitel werden die zentralen theoretischen Zugänge und Fragestellungen der Arbeit vorgestellt. Obwohl diese Grundlagen in den einzelnen Beiträgen jeweils separat diskutiert werden, ist die Verknüpfung journalismustheoretischer und techniksoziologischer Zugänge bei der Analyse der Rolle der automatisierten Textproduktion im Journalismus Klammer und Startpunkt für die Diskussion der Forschungsergebnisse in Kapitel IV. Die Arbeit folgt dabei übergeordnet dem Konzept der Ko-Evolution.

«Co-evolution is the linked evolution of two (or more) dynamics, each of which can be conceptualized in terms of variations and selections (and retentions). The linkage may give rise to patterns with dynamics of their own.» (Rip 2002: 10)

Dieses Verständnis erlaubt es, den Einfluss der automatisierten Textproduktion im Journalismus auf gleicher Ebene mit ökonomischen, politischen und kulturellen Triebkräften in technisch-sozialer Wechselwirkung zu analysieren (vgl. Dogruel 2013; Latzer 2013; Stewart & Williams 1998; Werle 2007). Dieser Wandel wird durch gegenseitigen selektiven Druck bzw. durch gegenseitige Anpassung vorangetrieben und ist durch ein «komplexes, d.h. adaptives, nicht-lineares Systemverhalten gekennzeichnet» (Latzer 2013: 241). Die technische Innovation «algorithmischer Journalismus» auf Basis der Natural Language Generation bildet dabei Kern und Triebkraft dieses Wandels und lässt sich als evolutionärer Prozess untersuchen (vgl. Dogruel 2013: 341f.). Diese in der Kommunikationswissenschaft noch unterrepräsentierte Perspektive überwindet die Dichotomie von technik- und sozial-deterministischen Interpretationen des Medienwandels und fokussiert dabei auf das Wechselspiel gesellschaftlicher Faktoren im Rahmen des Kontext-Modells von Weischenberg (1992) (vgl. Latzer 2013). Somit können Interaktionsprozesse und Wechselwirkungen auf gesellschaftlicher, organisatorischer und professioneller Ebene im Journalismus theoretisch und empirisch greifbar gemacht werden (vgl. Werle 2007: 38; 2005). Kombiniert wird die Analyse mit Ansätzen der Techniksoziologie und der Annahme, dass Technik nicht nur durch die Gesellschaft geprägt und geformt wird, sondern auch selbst aktiv als Akteur oder Institution sozial handeln kann und Beobachtungs- und Beschreibungsleistungen im Journalismus übernimmt (Dolata & Werle 2007; Just & Latzer 2016).

Wie u.a. Aitamurto und Lewis (2013), Lewis und Westlund (2015a,b), Young und Hermida (2015) oder auch Örnebring (2010) zeigen, profitiert insbesondere die Journalismusforschung von der Verknüpfung techniksoziologischer und journalismustheoretischer Zugänge bei der Analyse des Medienwandels. Dies ermöglicht nicht zuletzt auch die Offenheit der Journalismusforschung als Integrationsdisziplin an der Schnittstelle von Kultur-, Sozial- und Naturwissenschaften (vgl. Hanitzsch & Engesser 2014: 139; Karmasin 2003). Technik als Einflussfaktor in Bezug auf den Journalismus wird in diesen Arbeiten aber stark mit Bezug auf die Handlungsebene und die Actor-Network-Theorie von Latour

(1987) thematisiert (vgl. auch Pinch & Bijker 1984; Plesner 2009; Primo & Zago 2014; Schmitz-Weiss & Domingo 2010). Die Arbeit integriert deshalb – mit Bezug auf das Kontext-Modell von Weischenberg (1992) – nicht nur das techniksoziologische Verständnis von Technik als Akteur, sondern auch die Rolle von Technik als Institution in die Analyse.

Zunächst wird das Phänomen «algorithmischer Journalismus» im Forschungsprogramm «Algorithmische Selektion im Internet» verortet. Dieses Verständnis ermöglicht es, die vielfältigen Einflüsse von algorithmischen Anwendungen im Internet sichtbar zu machen (vgl. Latzer et al. 2014). Die Perspektive von Technik als Akteur und Institution erweitert den theoretischen Zugang zur Wirkungsweise von Technik. Unter Berücksichtigung dieser techniksoziologischen Ansätze wird das Kontext-Modell nach Weischenberg (1992) vorgestellt und mit den spezifischen Fragestellungen aus ethischer, technischer und ökonomischer Perspektive verknüpft.

1. Verortung im Forschungsprogramm «Algorithmische Selektion im Internet»

Um die Rolle und den Einfluss der automatisierten Textproduktion empirisch und theoretisch einzuordnen, wird mit Verweis auf Latzer et al. (2014) algorithmischer Journalismus als Anwendung der algorithmischen Selektion im Internet operationalisiert.

Anwendungen auf Basis algorithmischer Selektion gewinnen dabei mit der steigenden Bedeutung internetbasierter Informations- und Kommunikationstechnologien, der zunehmenden Datafizierung und ihrer Diffusion in verschiedene Lebensbereiche eine immer höhere gesellschaftliche und wirtschaftliche Relevanz (Anderson 2012; Gillespie 2014; Latzer et al. 2014; Mager 2012; Pasquale 2015). Latzer et al. (2014) kategorisieren diese Anwendungen als zentrale, strukturierende Innovationen sowie als Triebkräfte des aktuellen Medienwandels und der Informationsgesellschaft. Das Internet ermöglicht dabei die Entwicklung dieser Anwendungen und Innovationen aufgrund seiner spezifischen Architektur und Infrastruktur als komplexes, adaptives System (vgl. Just & Latzer 2016: 244)

Auf Basis empirischer Fallanalysen identifizieren Latzer et al. (2014) neun Gruppen von Internetanwendungen, die trotz divergierender Zwecke und Ziele automatisierte Selektions- und Relevanzentscheidungsprozesse als verbindende Elemente aufweisen und Einfluss auf gesellschaftliche Prozesse nehmen. Neben Such-, Überwachungs- oder Recommender-Anwendungen baue auch algorithmischer Journalismus auf diesem Grundprinzip auf. Seine Funktionalität wird analog zum Vorgehen der Autoren mit einem Eingabe-Verarbeitung-Ausgabe-(E-V-A-)Modell in Artikel II dargestellt. Im Rahmen dieser Operationalisierung ergeben sich für den algorithmischen Journalismus spezifische Herausforderungen, die in dieser Arbeit adressiert werden sollen.

Technisch zentral sind für alle Anwendungen des automatisierten Journalismus die Selektion (Eingabe), Gewichtung und Relevanzzuweisung strukturierter, digitaler Daten

durch Algorithmen (Verarbeitung). Die Zuweisung der Relevanz richtet sich dabei nach der Programmierung und der Beschaffenheit der Datensets. Die Daten werden also zunächst in eine semantische Struktur gebracht und dann in Textform in natürlicher Sprache über spezifische Kanäle auf einer reichweitenrelevanten Plattform durch einen professionellen journalistischen Akteur publiziert (Ausgabe). Nach initialer Programmierung kann das System relativ autonom, effizient (schnell) und meist sehr effektiv (fehlerfrei) journalistische Texte erstellen, publizieren und gegebenenfalls distribuieren. Artikel I wie auch Artikel II diskutieren algorithmischen Journalismus als Anwendung algorithmischer Selektion und analysieren seine spezifischen Leistungen im Rahmen des E-V-A-Modells.

Mit der steigenden Verfügbarkeit digitaler Daten, der Flut von Informationen und dem sich verändernden Leseverhalten erhöht diese Entwicklung einerseits den Bedarf an Reduktion von Komplexität durch den Journalismus, die Entwicklung ist aber gleichzeitig auch Basis und Triebkraft für die automatisierte Textproduktion. Denn die Verfügbarkeit strukturierter, digitaler Daten ist eine zentrale Voraussetzung dafür, Texte in großer Zahl automatisiert herzustellen. Aus koevolutionärer Sicht ist NLG – die technische Basis von algorithmischem Journalismus – damit eine Enabling Technology, da sie auf Grundlage ihrer technischen Spezifika Optionen für gesellschaftliche Anwendungen bereitstellt. Konkret zeigt sich das durch ihren Einsatz im Journalismus und durch die Übernahme spezifischer journalistischer Leistungen (vgl. Artikel II). Diese regelnde und steuernde Kraft gewinnt algorithmischer Journalismus aber nur in Kombination mit der Struktur und Architektur des Internets als Regelsystem und als Anwendung algorithmischer Selektion (vgl. Latzer et al. 2014; Lessig 1999; Just & Latzer 2016; siehe Abschnitt 2: *Technik als Institution*). Denn prinzipiell können Texte auch ohne Verknüpfung mit dem Internet automatisch erstellt werden, da es sich bei den meisten NLG-Systemen um geschlossene Software-Systeme handelt. Das Internet ist ein quasi unendlicher Datenpool und ermöglicht aufgrund seiner Architektur neben der Datensammlung auch die Distribution der automatisierten Texte an ein heterogenes Publikum. Zudem ist es auch Zugangspunkt für potentielle journalistische Kunden. Denn viele Anbieter bieten automatisierte Texte mittlerweile als Software as a Service (SaaS) an. Der Zugang zur Software ist cloudbasiert oder direkt über APIs in Content-Management-Systeme von Redaktionen eingebunden. Die spezifische Beschreibungs- und Ordnungsleistung von algorithmischem Journalismus entfaltet sich dann, wenn die automatisierten Texte auf Online-Plattformen journalistischer Akteure publiziert werden und der Leser Zugang zu diesen Inhalten hat, also Öffentlichkeit hergestellt wird (siehe auch Abschnitt 2: *Technik als Institution*).

In diesem Spannungsfeld ergeben sich Herausforderungen und Problemfelder auf unterschiedlichen Ebenen der Gesellschaft (für einen Überblick über Anwendungen algorithmischer Selektion siehe Saurwein et al. 2015). Auf ökonomischer Ebene stellen sich u.a. Fragen in Bezug auf Markt- und Geschäftsmodellentwicklungen im Journalismus. Mit dem

Einfluss der Technik ergeben sich beispielsweise auch ethische Fragen der Verantwortung, Manipulation und Transparenz (vgl. Artikel IV und V). Nicht zuletzt stellt sich die Frage, welche spezifischen Leistungen algorithmischer Journalismus bei der Beschreibung und Beobachtung der Gesellschaft übernehmen kann (vgl. Artikel II und III; zur algorithmischen Konstruktion von Realität vgl. Just & Latzer 2016)?

Die konkrete Nutzung von Anwendungen algorithmischer Selektion wird dabei im Wechselspiel mit politischen, ökonomischen und kulturellen Faktoren verhandelt (vgl. Latzer et al. 2014). Die inhaltliche Ausgestaltung dieses Prozesses kann jedoch je nach Anwendung und Einsatzzweck variieren (vgl. Just & Latzer 2016). Denn algorithmische Selektion hängt auch davon ab, welche Eingabeform verwendet wird. Oftmals wird Big Data auf dieser Ebene genutzt (vgl. Saurwein et al. 2015). Dieser Begriff wird in all seinen Facetten in der Literatur unterschiedlich operationalisiert (vgl. Wiesenberg et al. 2017: 2). Generell werden darunter ökonomische und gesellschaftliche Potentiale und Risiken beim Zugang sowie bei der Speicherung und Verarbeitung großer, unstrukturierter digitaler Daten aus internen und externen Quellen diskutiert (vgl. Flew et al. 2012; Mayer-Schönberger & Cukier 2013). Für eine Nutzung im algorithmischen Journalismus müssen diese digitalen Daten aber zunächst entweder manuell oder automatisiert strukturiert werden. Diese Strukturierung ist notwendig, damit die Software-Anwendungen die spezifische Eingabeaufforderung (Syntax, Code) bearbeiten können. Dies gilt sowohl für die Daten, die den Kern des algorithmischen Journalismus bilden, als auch für potentielle Datenbanken zu faktischem oder episodischem Wissen, die die Texte inhaltlich ergänzen können (vgl. Stede, Anhang A1). Dies ist eine zentrale Herausforderung für die automatisierte Texterstellung, denn einerseits sind die digitalen Daten aktuell nur eingeschränkt verfügbar, andererseits ist auch deren Qualität entscheidend. Nicht alle Daten haben einen journalistischen Wert, der den Programmierungs- und Kontrollaufwand einer skalierbaren Automatisierung rechtfertigen würde. Eine weitere Diffusion der automatisierten Textgenerierung hängt nicht zuletzt auch davon ab, ob Medienorganisationen in der Lage sind, diese Daten zu sammeln und zweckmäßig für bestimmte Themenbereiche einzusetzen. Dennoch ermöglicht die Architektur des Internets überhaupt erst, dass diese digitalen Daten generiert, gesammelt, verarbeitet und distribuiert werden können. Diese Entwicklung hat neben der finanziellen Unsicherheit und Schieflage im Journalismus nicht zuletzt dazu geführt, dass auf dem Gebiet der NLG in den letzten zehn Jahren massive Fortschritte gemacht wurden. Dabei gewinnt algorithmischer Journalismus vielmehr in Kombination mit anderen Anwendungen algorithmischer Selektion, wie z.B. Suchanwendungen oder Recommender-Systemen, an Bedeutung. Denn diese haben außerhalb der eigenen Plattform Einfluss darauf, wo die automatisierten Texte am Ende sichtbar werden. Als Ausprägung des Computational Journalism ist die automatisierte Textproduktion damit Teil des Journalismus und übernimmt Funktionen bei der Beobachtung und Beschreibung der Gesellschaft für

ausgewählte Themenbereiche (vgl. Artikel I und II). Sie reduziert Komplexität, indem Daten aus einer oftmals unübersichtlichen Struktur in Textform aufgearbeitet und damit kognitiv leichter zugänglich gemacht werden können. Auf der Ebene der professionellen Akteure reduziert sie Transaktionskosten und schafft Freiräume für andere journalistische Aufgaben, deren Kern nicht nur eine repetitive und faktenbasierte Berichterstattung ist. Die Diskussion dieser spezifischen Leistungen und der Einfluss der automatisierten Textproduktion im Journalismus werden dabei ausführlich in Kapitel IV diskutiert.

Das skizzierte Konzept der algorithmischen Selektion liefert den analytischen Überbau, um den Einfluss von algorithmischem Journalismus im Spannungsfeld der zunehmenden Algorithmisierung und Datafizierung der Gesellschaft zu verorten. Die Dissertation adressiert deshalb ausgewählte Fallanalysen zu ethischen, technischen, organisatorischen und ökonomischen Fragestellungen (für einen Überblick siehe Tabelle 1).

2. Technik als Akteur und Institution

Konzeptionelle Linse für die Analyse der Wirkungsweise von algorithmischem Journalismus ist in dieser Arbeit die Perspektive auf Technik als Akteur und Institution. Der Mehrwert der beiden Ansätze liegt in der Analyse der Wechselbeziehung von Journalist und Algorithmus und ermöglicht die Kombination von Fragen zur Wirkung von Technik mit einer kommunikationswissenschaftlichen Perspektive (siehe Abschnitt 3: *Das Kontext-Modell des Journalismus*). Diese beiden Ansätze identifizieren Technik als zentrale Triebkraft des Medienwandels, ohne deterministisch zu sein. Vor dem Hintergrund dieser theoretischen Überlegungen findet in Kapitel IV die Diskussion der Forschungsergebnisse der Dissertation statt.

Wie bereits erwähnt, ermöglicht die Integration einer techniksoziologischen Perspektive eine differenzierte Betrachtung der Strukturierungsleistungen von algorithmischem Journalismus. Die Komplexität des technischen Wandels und Fragen des Einflusses von Technik auf soziales Handeln können so in Kombination mit journalismustheoretischen Ansätzen reduziert und theoretisch wie empirisch greifbarer werden. Rammert (2007: 82ff.; auch Latzer 2013) verweist hier auf die steigende Bedeutung von Software-Agenten, die ihre Aufgaben relativ autonom, reaktiv, tätigkeitsbezogen und unter Bezugnahme auf andere Agenten ausführen. Mit dieser «Algorithmisierung» (Rammert 2007: 62) ist gemeint, dass Computer Daten verarbeiten, strukturieren, mit einer Relevanzzuweisung wieder kontextualisieren und – je nach Zweck der Anwendung – mit der Umwelt kompatibel machen. Für den Journalismus sind diese Prozesse in Bezug auf ihre originäre Leistung bei der Beobachtung und Beschreibung der Gesellschaft relevant. Denn die steigende Komplexität der Interaktion zwischen Mensch und Algorithmus erhöht zwangsläufig

fig auch die an Medienmacher und Medienbesitzer zu stellenden Anforderungen. Sie müssen in diesem Wechselspiel die zentralen Leistungen des Journalismus garantieren (vgl. Artikel IV). Diese werden in Artikel II identifiziert. Erste Forschungsergebnisse zeigen, dass Internetanwendungen auf Basis algorithmischer Selektion massenmediale Aufgaben übernehmen können und damit direkt an der gesellschaftlichen Konstruktion von Wirklichkeit beteiligt sind (vgl. Beer 2009; Just & Latzer 2016: 10f.).

Daher sollen technische Prozesse in der Gesellschaft hier nicht als «(s)innfremde Vorgänge und Gegenstände» untersucht werden, die nur als «Anlaß, Ergebnis, Förderung oder Hemmung menschlichen Handelns in Betracht» kommen (Weber 1972 [1922]: 13; zitiert nach Schulz-Schaeffer 2007: 2), Technik kann auch selbst strukturbildend sein oder als Akteur handeln (vgl. Dolata 2011: 46ff.). Technische Innovationen können damit nicht nur als wissenschaftlich-technische Neuerungen, sondern auch als organisatorische, politische, kulturelle und institutionelle Veränderungen verstanden werden (vgl. Werle 2007: 31). Dabei lassen sich grundsätzlich drei Ansätze unterscheiden. Technik kann als eigenständiger Handlungsträger (Akteur), als Institution (Regel) oder als Struktur (System) beschrieben werden (vgl. Dolata & Werle 2007: 17; Dolata 2011). Akteurskonzepte von Technik – z.B. die Actor-Network Theory von Bruno Latour (1987) oder das Konzept der gradualisierten Handlungsträgerschaft von Technik (vgl. Rammert & Schulz-Schaeffer 2002; Rammert 2007: 109ff.) – setzen auf der Mikroebene an. System- und Institutionenkonzepte greifen Fragestellungen zur Wirkung von Technik auf, die eher auf der Meso- oder der Makroebene zu verorten sind (vgl. Winner 1986; Lessig 1999).

Mit Fokus auf die Perspektive von Technik als Akteur und Institution kann der kommunikationswissenschaftliche Ansatz erweitert werden, der die Ausprägung und Gestaltung von Technik meist nur im Rahmen von Organisationen und Institutionen betrachtet (vgl. Neverla 2003). Dieses übergeordnete Verständnis der technisch-sozialen Wirkung kann als Ko-Evolution von Technik und Gesellschaft gefasst werden (vgl. Latzer 2013: 237ff.; Werle 2007). Algorithmischer Journalismus wird dabei als Triebkraft im Medienwandel identifiziert und analysiert.

Einerseits wird der Begriff «Technik» in der Literatur unterschiedlich konzeptualisiert (enger und weiter Technikbegriff), andererseits ist auch die Abgrenzung vom Begriff «Technologie» nicht immer trennscharf (vgl. Häußling 2014: 117ff.; Schulz-Schaeffer 2008). Dies oftmals auch deshalb, weil im Englischen auf eine solche Unterscheidung verzichtet und der Begriff «technology» übergreifend verwendet wird. In Anlehnung an Rophol (1991: 22) und Weischenberg (1995: 15) bezieht sich der Begriff «Technik» in dieser Arbeit pragmatisch auf Werkzeuge und Verfahren, der Begriff «Technologie» auf Klassen von technischen Werkzeugen, die wissenschaftlich-systematisch entwickelt wurden. Natural Language Generation – die technische Basis für algorithmischen Journalismus – ist damit eine zentrale Technik der Sprachtechnologie in der Computerlinguistik.

Es ist nicht das Ziel dieser Synopse, einen breiten Überblick über die Rolle von Technik in der industrie- und techniksoziologischen Literatur zu vermitteln. Vielmehr soll an dieser Stelle und im Rahmen des koevolutionären Verständnisses dieser Arbeit gezeigt werden, welchen Beitrag der techniksoziologische Zugang und die Integration der Rolle von Technik als Akteur und Institution für die Analyse von algorithmischem Journalismus leisten und welche Anknüpfungspunkte sich damit für die theoretische Diskussion und für die Verortung der Einflusshierarchien im Kontext-Modell ergeben. Dabei wird mit einem Schwerpunkt auf der Rolle von Technik als Akteur auch die Konzeption von Technik als Institution kurz skizziert.

Technik als Akteur

Ein Schwerpunkt der Betrachtung von algorithmischem Journalismus soll hier auf der Konzeption von Technik als Akteur auf der Handlungsebene und – mit Verweis auf das Kontext-Modell – auf der professionellen Ebene liegen. Zwar monieren einige Kritiker, dass handlungstheoretische Ansätze nicht in der Lage seien, Phänomene auf der Meso- und Makroebene zu erklären (vgl. Altmeyden 2000; Dolata 2011: 47; Weißer 2002), doch die Akteursfrage spielt gerade bei der Analyse des Wechselspiels zwischen Mensch und Technik im algorithmischen Journalismus eine entscheidende Rolle. Einerseits stellt sich die Frage, welche spezifischen Leistungen Algorithmen hier im Austausch mit den Journalisten übernehmen können (vgl. Artikel I, II, III & IV). In Artikel I wird vor dem Hintergrund der Digitalisierung und Datafizierung algorithmischer Journalismus im Forschungsfeld im Computational Journalism verortet und mit Verweis auf die Rolle der Daten, die Rolle des Algorithmus und die Rolle des Journalisten in der Nachrichtenproduktion differenziert. Die technische Funktionsweise und das journalistische Potential wird in Artikel II diskutiert. In Artikel III wird das Wechselspiel sichtbar gemacht und zudem von professionellen Journalisten bewertet. Andererseits spielt die Rolle von Algorithmen als Akteur auch bezüglich ethischer Fragen eine Rolle, die Artikel IV und V aufgreifen, nämlich die Fragen, ob und in welchem Ausmaß Technik moralisch handeln und als Träger von Verantwortung gelten kann (vgl. Kroes & Verbeek 2014). Zudem beeinflusst das Wechselspiel zwischen Journalist und Technik auch die organisatorische Ebene, z.B. in Bezug auf Fragen der Redaktionsorganisation oder der Entwicklung neuer Märkte (siehe dazu den Abschnitt *Technik als Institution*). Dabei geht es nicht darum, Technik «Intentionen und damit Bewusstsein bzw. Bewusstsein in ihrem Operieren zu unterstellen oder sie zu einem Subjekt, das in und aus Freiheit handelt, zu stilisieren» (Häußling 2014: 95). Es geht vielmehr darum, Technik in Wechselbeziehung zu menschlichen Akteuren und damit als Teil gemeinschaftlicher Tätigkeiten zu untersuchen (vgl. Rip 2007; Neuberger 2009; Latzer 2013). Handlungsträgerschaft von Technik ist hier als mehrstufiger Prozess zu verstehen. Denn das Potential von algorithmischem Journalismus kann erst durch das Zusammenspiel verschiedener Relevanz- und Selektionsleistungen abgerufen werden. Zum Beispiel

dann, wenn die Distribution automatisierter Inhalte – beispielsweise über Facebook – von plattforminternen Filter- und Empfehlungs-Algorithmen abhängig ist und so der Zugang zum Leser außerhalb der Einflussosphäre des Journalisten oder der Organisation liegt.

Ein zentraler Streitpunkt der Diskussion zur Wirkungsweise von Technik ist, ob technische Artefakte selbst als Entitäten zu betrachten sind und ihnen so im Sinne der soziologischen Theorie Handlungsträgerschaft (Agency) zugeschrieben werden kann (vgl. Dolata 2011: 46ff.). Die Diskussion innerhalb der Agency-Debatte kann dabei in vier Bereiche gegliedert werden (vgl. Matsuzaki 2011: 301ff.): Zuerst sind Konzepte zu nennen, die ein intentionales Handlungsverständnis in den Mittelpunkt stellen und den Akteursstatus von Technik soziologisch mit dem Unterschied zwischen Mensch und Maschine aberkennen (vgl. Collins & Kusch 1998). Zweitens die Ansätze von Bruno Latour (1987), in denen der Handlungsbegriff auf den Aspekt der Veränderung reduziert wird und Mensch und Nichtmensch gleichgesetzt werden. Drittens die Ansätze, die als unmittelbare Kritik an dieser Operationalisierung eine Differenzierung des Handlungsbegriffs vornehmen (vgl. Rammert & Schulz-Schaeffer 2002) und viertens Ansätze, die Technik als fiktiven Akteur betrachten (vgl. Werle 2002).

Die Arbeit fokussiert sich auf das Konzept des gradualisierten Handelns von Werner Rammert und Ingo Schulz-Schaeffer (2002), da dieses sich gut eignet, um die Interaktionsebenen zwischen Journalist und Algorithmus aufzuzeigen (zum Interaktions- und Interaktivitätsbegriff in der Kommunikationswissenschaft Neuberger 2007). Denn dieses Konzept ermöglicht die Diskussion über den Handlungsrahmen – das Niveau, die Koppelung und Zurechnung von Handlungsträgerschaft – von Algorithmen jeweils in Abhängigkeit vom Stand der technischen Entwicklung sowie der Art und Weise der gesellschaftlichen Thematisierung und ermöglicht so Implikationen für die gesellschaftliche, organisatorische und professionelle Ebene des Journalismus (vgl. Rammert & Schulz-Schaeffer 2002: 43ff.; Rammert 2007).

Folgt man den Autoren, besteht die einfachste Form des Handelns darin, einen Unterschied herzustellen (*Level 1: «mechanisches Bewirken-Können» i.S.v. Kausalität*). Als zweite Form wird ein Aktionsvollzug unterschieden, bei dem der Akteur einen Spielraum hat, der es ihm ermöglicht, entweder auf diese oder auf jene Weise eine Wirkung zu entfalten (*Level 2: «Auch-anders-handeln-Können» i.S.v. Kontingenz*). Davon unterscheiden die Autoren noch eine komplexere Form des Handelns, bei der die Akteure ihre Handlungen reflektieren und sinnhaft begründen können (*Level 3: «intentionales Verhalten» i.S.v. Intentionalität*) (vgl. Rammert & Schulz-Schaeffer 2002: 49f.). Welche der drei Formen dieses Handlungsbegriffs angewendet werden, hängt vom jeweiligen Interaktionszusammenhang zwischen Mensch und Maschine ab. Die Dreiteilung des Handlungsbegriffs ist insofern konsistent, als das Handeln von Maschinen menschlichen Akteuren ähnlich ist, wenn man lediglich den Ak-

tionsvollzug betrachtet (vgl. Rammert & Schulz-Schaeffer 2002: 43ff.). Innerhalb dieses vorgestellten Handlungsbegriffs gibt es unterschiedliche Freiheitsgrade der Handlungsträgerschaft. Auf der Ebene des Auch-anders-handeln-Könnens (Level 2) reicht das Spektrum von der Wahl vordefinierter Alternativen in Software-Programmen bis hin zur vergleichsweise «freien» Selbstgenerierung von lernfähigen Algorithmen (vgl. Rammert & Schulz-Schaeffer 2002: 50).

Mit Verweis auf die Strukturierung des Forschungsfeldes in Artikel I und die technische Potentialanalyse in Artikel II sowie die Darstellung der technischen Funktionsweise im Rahmen des Eingabe-Verarbeitung-Ausgabe-Modells kann so das Wechselspiel von Journalist und Algorithmus interpretiert und der Anteil an der Leistungserstellung diskutiert werden. Diese Ausdifferenzierung ist speziell auf der professionellen Ebene des Kontext-Modells zu diskutieren. Der Einfluss von externen Akteuren wie z.B. den Programmierern kann damit aber nur indirekt durch den Code bzw. durch die Texte selbst sichtbar werden. Der Ansatz der gradualisierten Handlungsträgerschaft ermöglicht zudem eine differenzierte Betrachtung der ethischen Herausforderungen, da sich die Einflussebenen und Interaktivitätslevel zwischen Journalisten und Algorithmus schematisch darstellen lassen (vgl. Artikel IV und V). Fragen zur Verantwortung, Transparenz und Kontrolle können so strukturiert und diskutiert werden. Die spezifischen Forschungsergebnisse der Dissertation werden mit Verweis auf diese Operationalisierung in Kapitel IV diskutiert.

Technik als Institution

Um die Nachteile der Mikrofokussierung auszugleichen und die Wirkungsweise von Technik auch auf der Meso- und der Makroebene einzubeziehen, wird die Analyse von algorithmischen Journalismus mit der Perspektive von Technik als Institution kombiniert. Denn insbesondere der Ordnungscharakter von Technik wird in der Medientechnik (vgl. Katzenbach 2012) und bei algorithmischen Prozessen im Internet sichtbar und eignet sich als weitere Untersuchungsebene (vgl. Napoli 2014; Just & Latzer 2016). An dieser Stelle soll das Potential dieses Ansatzes für die finale Diskussion in Kapitel IV skizziert werden.

Das Konzept von Technik als Institution betont dabei die prägende Wirkung der Technik auf gesellschaftliche Prozesse. Sie hat eine regelsetzende und steuernde Funktion für Arbeitsprozesse und für die Organisationsstruktur durch technische Codes und Normen (vgl. Dolata & Werle 2007; Latzer 2013). Diese Normen regeln aber nicht nur die Funktionsweise und Leistungen von algorithmischem Journalismus (vgl. Artikel II), sondern auch den Umgang mit diesem (vgl. Artikel III). Sie rufen verschiedene «industrielle, politische und gesellschaftliche Möglichkeiten ihrer Organisation, Regulierung und Nutzung» (Dolata 2011: 49) hervor, die im Verlauf der technischen und gesellschaftlichen Entwicklung verändert und angepasst werden können (vgl. Schmidt & Werle 1998: 108ff.). Auch

dem Journalismus wird eine institutionelle Kraft zugeschrieben, indem er eine Orientierungsfunktion für soziales Handeln übernimmt und als soziales Ordnungsgebilde fungiert, das sozial geteilte Regeln hervorbringt (vgl. Brüsemeister 2008: 24; Neuberger 2013; Kiefer 2010; Scholl 1997: 471f.). Dabei führt der Einsatz der automatisierten Textgenerierung im Journalismus nicht nur zu ökonomischen Fragestellungen, beispielsweise zu sich verändernden Marktbeziehungen und Nachfragemustern (vgl. Artikel II), zu organisationalen Herausforderungen auf der Ebene der Redaktionsorganisation und der Arbeitsabläufe, sondern auch zu ethischen Fragestellungen der Transparenz, Objektivität, Rechtmäßigkeit und damit auch zu möglichen Fragen der Regulierung (vgl. Artikel IV und V). Müssen automatisierte Inhalte als solche gekennzeichnet werden oder gar auf organisatorisch-institutioneller Ebene bestehende Ethik-Kodizes zum Umgang mit algorithmischem Journalismus angepasst oder entwickelt werden?

Wie bereits skizziert, ist algorithmischer Journalismus maßgeblich an die Struktur und Architektur des Internets als Regelsystem gebunden. Dadurch gewinnt die automatisierte Textproduktion eine regelnde und steuernde Kraft und kann im Sinne Lessigs (1999) und seiner Analyse von intelligenter Informationstechnik («Code is law») bezüglich ihrer Funktion mit sozialen Regelsystemen gleichgesetzt werden (vgl. Latzer et al. 2014; Just & Latzer 2016; siehe auch Abschnitt 1: *Verortung im Forschungsprogramm «Algorithmische Selektion im Internet»*).

Die Analyse der Wirkungsweise von Technik als Akteur und Institution hilft die Rolle von algorithmischem Journalismus auf der Mikro-, Meso- und Makroebene zu verorten. Diese Strukturierung eignet sich, um Fragestellungen zur Wirkungsweise von Technik analog mit einer kommunikationswissenschaftlichen Perspektive und dem Kontext-Modell von Weischenberg (1992) auf gesellschaftlicher, organisatorischer und professioneller Ebene zu verknüpfen.

3. Das Kontext-Modell des Journalismus

Für die Analyse des Einflusses der automatisierten Textproduktion auf den Journalismus greift die Arbeit auf das Kontext-Modell von Weischenberg (1992) zurück. Mit diesem Modell ist es möglich, die vielfältigen und komplexen Einflussebenen und -hierarchien im Journalismus zu klassifizieren und den Einfluss von Technik für eine empirische Analyse sichtbar und zugänglich zu machen. Das Modell – entworfen für Gesellschaften des westlichen Typs – hat sich dabei auch in anderen Arbeiten als Analyserahmen heuristisch bewährt (vgl. Scholl & Weischenberg 1998: 216). Je nach untersuchungsleitender Fragestellung wurde es teilweise modifiziert und erweitert (vgl. u.a. Dulinski 2003; Esser 1998; Fabris 2000; Hanitzsch 2009; Redelfs 1996; Scholl & Weischenberg 1998; Schulten-Jaspers

2013). Da der Diskurs darüber, welche Funktionen und Aufgaben Journalismus übernehmen muss und kann, sehr vielfältig ist, ermöglicht ein integratives Verständnis von Journalismus die Einflüsse von Technik im Spannungsfeld der Digitalisierung sichtbar zu machen und entlang dieser Ebenen zu diskutieren (vgl. Neuberger 2002, 2009). Dennoch kritisieren u.a. Baum (1994), Löffelholz (2000: 50), Neuberger (1996: 185ff.) und Esser (1998) das Modell im Hinblick auf theoretische Brüche in seiner systemtheoretischen Konzeption, eine mangelnde Berücksichtigung der Rolle des Akteurs sowie die Missachtung potentieller Wechselwirkungen zwischen den Ebenen.^{4,5} Diese Arbeit geht in ihren Beschreibungs- und Erklärungsversuchen daher von einem «Nebeneinander von Kommunikation und Handlung, von System und Akteur, von Autonomie und Fremdsteuerung» (Hohlfeld 2003: 107) aus. Für die Analyse von algorithmischem Journalismus ist das Kontext-Modell von Weischenberg (1992) im Sinne Essers (1998: 26) als integratives Modell und als heuristischer Analyserahmen zu verstehen. Die Ebenen des Kontext-Modells stehen in einem Interaktionsverhältnis und beeinflussen sich gegenseitig. Jede Variable gewinnt ihren Einfluss im Verbund und in Wechselwirkung mit den anderen Einflussfaktoren der gesellschaftlichen, organisatorischen und professionellen Ebene. Diese Strukturierung dient dabei in erster Linie der systematischen Analyse der Rolle von Technik sowie ihres Einflusses auf den Journalismus und ist eingebettet in das koevolutionäre Verständnis von medialem Wandel (vgl. Latzer 2013). Die Interdependenzen können im Rahmen des Modells diskutiert und verortet werden. Die Analyse des Wandels im Journalismus ist damit im Sinne Blöbaums (2005) auch als eine Untersuchung der Veränderung journalistischer Strukturelemente zu sehen. Die Verknüpfung von Makro-, Meso- und Mikroperspektive ist zudem eine zentrale Voraussetzung, um handlungsprägende sowie strukturbildende Einflussfaktoren zu identifizieren und zu analysieren (vgl. Wyss & Keel 2010: 341).

Das Kontext-Modell von Weischenberg (1992) besteht aus vier hierarchischen Ebenen und hilft den Einfluss von algorithmischem Journalismus auf das journalistische Handeln zu verorten. Die Ebene der Medienakteure (Rollenkontext, professionelle Ebene) bilden die Journalisten mit ihren Merkmalen, Einstellungen, Rollenselbstverständnissen und Sozialisierungen. Sie sind dabei Teil des Funktionskontexts der Medienaussagen und des Strukturkontexts der Medieninstitutionen (ökonomische, politische, organisatorische und

4 Weischenberg selbst (2014: 159ff.) präzisiert sein Modell gerade im Hinblick auf diese Kritik und stellt fest, dass das Modell eine «offene Beobachtungsperspektive» anbiete. «Für diese (zeitabhängige) Beobachtung des Journalismus bietet das Modell mit den Kontexten und Kategorien Orientierungshilfen an» (Weischenberg 2014: 162).

5 In englischsprachigen Arbeiten wird oftmals das ähnlich konzeptualisierte «Hierarchy-of-Influences»-Modell von Shoemaker und Reese (1996) herangezogen. Die Autoren unterscheiden fünf Einflussphasen: die Ebene der Individuen, die Ebene der professionellen Routinen, die Ebene der Organisationen, die außermediale Ebene und die ideologische Ebene. Weitere Modelle für eine Systematisierung von Einflusshierarchien im Journalismus liefern u.a. das Sphären-Modell von Donsbach (1987) sowie auch Esser (1998), McQuail (2000: 248), Ettema et al. (1987), Whitney et al. (2004) und Preston (2009).

technologische Imperative auf der organisatorischen Ebene). Mit dem Mediensystem als äußerem Kreis (Normenkontext; gesellschaftliche Ebene) und der Berücksichtigung gesellschaftlicher Rahmenbedingungen, historischer und rechtlicher Grundlagen, der Kommunikationspolitik sowie professioneller und ethischer Standards schließt das Modell ab (vgl. Weischenberg 1992: 69f.; Weischenberg et al. 2006a,b). Wenngleich der empirische und theoretische Teil der Artikel I bis V nicht explizit die Ebene der Medienaussagen umfasst, wird der Einfluss von algorithmischem Journalismus auf diese Ebene in Relation zu den anderen Ebenen gesetzt und diskutiert. Denn «Normen, Strukturen, Funktionen und Rollen bestimmen in einem Mediensystem, was Journalismus ist, der dann nach diesen Bedingungen und Regeln Wirklichkeitsentwürfe liefert» (Weischenberg 1992: 67).

Im folgenden Abschnitt werden die Ebenen des Kontext-Modells vorgestellt und im Hinblick auf die spezifischen Fragestellungen der Artikel I bis V diskutiert. Es wird gezeigt, welche Einflussvariablen in der Dissertation diskutiert werden und an welche theoretischen Konzepte diese anknüpfen.

3.1 Ebene der Mediensysteme

Auf gesellschaftlicher Ebene (Mediensysteme: Normenkontext) beobachtet Journalismus andere Teilbereiche der Gesellschaft und stellt der Öffentlichkeit gesellschaftlich relevante, nichtfiktionale Themen und Aussagen mit einer bestimmten Relevanz zur Verfügung, die nach professionellen und ethischen Standards innerhalb der gesellschaftlichen Rahmenordnung produziert werden (Weischenberg 1992: 68; Weischenberg et al. 2006b: 346f.). Auf dieser Ebene werden auch die allgemeinen Zielvorgaben für die Qualität von Medienprodukten formuliert, die auf alle Ebenen, insbesondere auch auf die Ebene der Medienaussagen, einwirken. Dazu gehören u.a. rechtliche Bestimmungen sowie «allgemeine und spezifische Werte und Normen, die im Rahmen der ethischen Diskurse und damit auch der Qualitätsdiskurse eine Rolle spielen» (Weischenberg 2014: 89; 2006; Arnold 2008).

Journalismus hat folglich eine gesellschaftliche und öffentliche Aufgabe inne, indem er Nachrichteninhalte unter Rückbezug auf professionsethische Kriterien selektiert, aufbereitet und veröffentlicht (vgl. Brosda 2010: 259). Vor dem Hintergrund dieser Leistungen und Aufgaben des Journalismus verortet die Arbeit die ethischen Herausforderungen des algorithmischen Journalismus aber nicht nur auf gesellschaftlicher und organisatorischer Ebene mit ihren strukturellen Rahmenbedingungen des journalistischen Handlungsfeldes (Institutionenethik), sondern bezieht mit Verweis auf das Kontext-Modell von Weischen-

berg (1992) und Pürer (1992) auch die professioneller Ebene im Hinblick auf eine individuelle Ethik des Journalisten in die Analyse ein (vgl. Artikel V).⁶ Denn es ist notwendig, nicht nur den «handelnde[n] Journalist[en] zum Ausgangspunkt der Betrachtung [zu] machen» (Thomaß 2003: 159), sondern auch das «Zusammenwirken unterschiedlicher Strukturen, in denen vielfältige Werte, Normen, Rollen, Stellen, Techniken usw. [...]» (Thomaß 2016: 541) zu berücksichtigen und in die ethische Analyse einzubeziehen (vgl. auch Köberer 2014). «Nicht mehr der einzelne als ‹ganzer Mensch› macht Journalismus, sondern Journalismus wird durch organisatorisches Handeln produziert» (Rühl 1996: 93f.). Diese integrierende Betrachtungsweise unterstreicht wiederum das koevolutionäre Verständnis der Arbeit und verweist zudem auf das Wechselspiel der Ebenen im Kontext-Modell.

Algorithmischer Journalismus setzt professionsethisches Handeln unter Druck, da Handlungsorientierungen und -anweisungen für den Umgang mit und die Produktion von journalistischen Inhalten auf organisatorischer wie individueller Ebene kontinuierlich angepasst werden müssen (vgl. Artikel IV). Um diese spezifischen Handlungsempfehlungen leisten zu können, müssen die unterschiedlichen Handlungsfelder im algorithmischen Journalismus aus normativer Perspektive analysiert werden. Die Analyse der Funktionsweise von algorithmischem Journalismus in Artikel II sowie auch die praktische Arbeit der Journalisten mit der Software in Artikel III liefern damit die empirische Basis für die theoretische Diskussion und Verortung der ethischen Herausforderungen in Artikel IV und V.

Für die Auseinandersetzung mit normativen Fragestellungen im journalistischen Handlungsfeld lässt sich die journalistische Ethik sinnvoll als theoretischer Bezugspunkt heranziehen. Die journalistische Ethik wird in dieser Arbeit als Teilbereich der Medienethik und als normative Ethik verstanden, die wiederum als Teildisziplin der philosophischen Ethik der praktischen Philosophie zugeordnet ist.⁷ Die journalistische Ethik hat dabei die Aufgabe, «Normen und Werte journalistischen Handelns unter Rückbezug auf Theorien der allgemeinen Ethik zu begründen, journalistische Ideale in Bezug auf die Möglichkeiten ihrer Realisierbarkeit zu überprüfen und normative Handlungsorientierung bereitzustellen» (Köberer 2014: 67; Rath 2006). Das hier konzeptionierte Verständnis von journalistischer Ethik berücksichtigt dabei sowohl die strukturellen Rahmenbedingungen als

⁶ Es existieren verschiedene Systematisierungsvorschläge für die journalistische Ethik. Für einen Überblick über systemtheoretische Annahmen und diskursethische Ansätze siehe die Arbeiten und Überblickswerke von Debatin & Funiok (2003) und Brosda & Schicha (2010).

⁷ Als angewandte Ethik hat die Medienethik die Aufgabe, Wert- und Normvorstellungen für mediales Handeln systematisch zu begründen und die Verknüpfung von Theorie und Praxis unter Rückbezug auf ethisch fundierte, normative Maßstäbe zu ermöglichen (vgl. Köberer 2014: 25). Sie untersucht Strukturen und Bedingungen des medialen Handlungsfeldes und beurteilt dabei auch Rezeptions- und Wirkungsweisen der Medieninhalte nach normativen Kriterien die universale Gültigkeit beanspruchen können. Um überhaupt Handlungsempfehlungen für die journalistische Praxis geben zu können, braucht die Medienethik Informationen über den aktuellen Stand technischer Entwicklungen (vgl. Köberer 2014: 34; Rath 2006).

auch die Rollen- und Aufgabenbezogenheit der Journalisten im journalistischen Handlungsfeld. Ausgehend von der Potentialanalyse und der Journalistenbefragung in Artikel II und III identifizieren Artikel IV und V die ethischen Herausforderungen im Umgang mit algorithmischem Journalismus. Hier stellt sich die Frage, ob und wie professionsethische Kriterien der Nachrichtenproduktion auch auf den algorithmischen Journalismus angewendet werden müssen, z.B. bei der Kennzeichnung der Quellen (Daten und Software). Zudem erlaubt die Potentialanalyse in Artikel II auch technische Gegebenheiten und Verantwortungsebenen der Nachrichtenproduktion in die ethische Diskussion aufzunehmen. Dieses multiperspektivische und -methodische Vorgehen auf gesellschaftlicher, organisatorischer und professioneller Ebene ermöglicht es erstmals, ein breites Bild der spezifischen ethischen Herausforderungen im algorithmischen Journalismus zu zeichnen.

Im Rahmen dieser Diskussion der journalistischen Ethik bezieht sich die normative Beurteilung journalistischen Handelns immer auch auf die Sicherung journalistischer Qualität (vgl. Artikel IV). Dabei wird die ethische Diskussion im Journalismus zunehmend durch die Diskussion über journalistische Qualität beeinflusst (vgl. Arnold 2008; Haller 2010; Fabris & Renger 2003; Thomaß 2016). Da es den einen verbindlichen Qualitätsmaßstab nicht gibt, geht es bei der Bestimmung medialer Qualität immer auch um die jeweils spezifisch zu thematisierenden und offenzulegenden Kriterien, Ziele und Inhalte des Begriffs (vgl. Karmasin 2005: 21; Köberer 2014; Artikel IV). Mit Bezug auf das Kontext-Modell und die normative Diskussion von Qualität ist in dieser Arbeit die Ebene des Mediensystems relevant, da hier allgemeine und normative Zielvorgaben für die Qualität der Produkte formuliert werden (vgl. Weischenberg 2006b: 669). Wie Weischenberg (2006a,b) und Altmeyden (2003b: 114) feststellen, ist die Analyse der Medienqualität aber an vielfältige Strukturelemente gebunden, die erst in einem institutionellen Kontext und im Wechselspiel der vier Ebenen des Kontext-Modells wirksam werden. Dies gilt auch für die Analyse der normativen Qualität im Rahmen der journalistischen Ethik. Deshalb werden Prozesse der Qualitätssicherung auf der Ebene der Institutionen, speziell in Medienorganisationen und im Redaktionsmanagement, der Einfluss auf die Ebene der Mediaussagen (Maßstäbe für Qualität; Messung von Qualität der Medienprodukte) sowie auch die Einstellungen der Journalisten gegenüber professionellen Standards und Arbeitsmethoden sowie Produkten auf der professionellen Ebene in den Artikeln III, IV und V mitdiskutiert (vgl. Weischenberg 2006a,b; Weischenberg et al. 2006a,b).

Diese Differenzierung ermöglicht erstmals auch eine übergeordnete Analyse der Verteilung und Zuschreibung von Verantwortung im algorithmischen Journalismus aus einer individualethischen (Akteursethik) und sozialetischen Perspektive (Institutionenethik) in Artikel IV und V (Pürer 1992; Weischenberg 1992: 170ff.). Auch Einflüsse der Technikethik werden in Artikel V im Rahmen der journalistischen Ethik diskutiert (vgl.

Maner 1996; Spinello 2011). Transparenz als Wert medialer Qualität und zentrales Kriterium normativer Fragestellung wird dabei exemplarisch für den algorithmischen Journalismus in Artikel IV beleuchtet.

In diesem theoretischen Kontext beleuchtet, verortet und diskutiert die Dissertation erstmals die Herausforderungen im algorithmischen Journalismus aus journalismusethischer Perspektive in Artikel III, IV und V:

- a) Welche Anforderungen stellen die Digitalisierung und die Verfügbarkeit von digitalen Daten im Journalismus an professionsethisches Handeln?*
- b) Wo und wie lassen sich ethische Herausforderungen an algorithmischen Journalismus theoretisch verorten?*
- c) Welche ethischen Herausforderungen für algorithmischen Journalismus lassen sich auf professioneller, organisatorischer und gesellschaftlicher Ebene identifizieren?*
- d) Inwieweit erfüllt algorithmischer Journalismus normative Qualitätsansprüche der Verantwortung und Transparenz?*
- e) Wie schätzen Journalisten die ethischen Chancen und Risiken von algorithmischem Journalismus aus professioneller Perspektive ein?*

3.2 Ebene der Medieninstitutionen

Die organisatorische Ebene (Medieninstitutionen: Strukturkontext) umfasst die Produktion der Aussagen in Medieninstitutionen nach spezifischen Regeln und Werten und schließt ökonomische, politische, organisatorische und technologische Imperative ein (vgl. Weischenberg 1992: 237ff.). Diese Imperative entscheiden bis zu einem gewissen Grad, «welche Wirklichkeitsentwürfe von Medien jeweils angeboten werden» (Weischenberg 1992: 237). Daher ist es wichtig, diese Faktoren bei der institutionellen Analyse von algorithmischem Journalismus einzubeziehen. Die Arbeit konzentriert sich dabei auf die Diskussion ökonomischer, organisatorischer und technologischer Imperative der automatisierten Textproduktion für den Journalismus. Nachfolgend werden die theoretischen Grundlagen für ihre Analyse auf organisatorischer Ebene erläutert. Für algorithmischen Journalismus ergeben sich u.a. folgende Fragestellungen, die im Rahmen dieser Arbeit in Artikel II und III berücksichtigt werden:

- a) Wer sind die zentralen Software-Anbieter von algorithmischem Journalismus?*
- b) In welcher Marktphase befindet sich der algorithmische Journalismus?*
- c) Hat Natural Language Generation (NLG) das technische Potential, traditionelle Funktionen des Journalismus in der Nachrichtenberichterstattung zu übernehmen?*
- d) Wie schätzen Journalisten die technischen und ökonomischen Chancen und Risiken von algorithmischem Journalismus aus professioneller Perspektive ein?*

Ökonomische Imperative

Die Medienökonomie analysiert die Bedingungen journalistischer Produktion, der Distribution und des Konsums von Medieninhalten und Trägermedien. Für den Medienbereich kann man «Ökonomisierung» damit als wachsende Relevanz ökonomischer Prinzipien und Handlungsrationitäten bei der «Institutionalisierung, Diversifizierung, Produktion und Konsumtion von Medien bzw. deren Inhalten» (Meier & Jarren 2001: 146) definieren. Renger (2006: 165) stellt fest, dass das «Gedeihen» des Journalismus heute weitgehend einem an Effizienz orientierten Wachstum gewichen sei. Auch der algorithmische Journalismus kann aufgrund seines technischen Potentials auf diese Ebene einwirken. Denn die gesellschaftliche Bedeutung des algorithmischen Journalismus als Anwendung algorithmischer Selektion ergibt sich auch durch seine Diffusion. Latzer et al. (2016) zeigen mit Rückgriff auf Ansätze der Netzwerk-Ökonomie, dass Anwendungen auf Basis algorithmischer Selektion einen ökonomischen Einfluss auf entstehende, aber auch auf bereits etablierte Märkte haben können (siehe auch Abschnitt 2: *Technik als Institution*). Dabei ist es möglich, die Ursachen für eine zunehmende Ökonomisierung – den Zusammenhang zwischen Angebot und Struktur von Medienmärkten – mit dem industrieökonomischen Marktstruktur-Marktverhalten-Marktergebnis-Modell (SVE) zu operationalisieren und zu analysieren (vgl. Bain 1956; Mason 1949). Das SVE-Paradigma wird hier als strukturierender Analyse-rahmen verstanden, der Rück- und Wechselwirkungen innerhalb des Modells berücksichtigt (vgl. Czygan & Kallfaß 2003: 293; Haas & Wallner 2008).

Der zentrale Beitrag der Dissertation liegt hier in der Diskussion des ökonomischen Potentials der Technik aus der Perspektive von Journalisten in Artikel III sowie in der Strukturierung des Anbietermarkts in Artikel II. Mit der Analyse und Einordnung von algorithmischem Journalismus in Marktphasen ist es möglich, Rückschlüsse auf das Wachstum des spezifischen Marktes der Texterstellung und damit auch Hinweise auf eventuelle Veränderungen des Medienmarkts zu gewinnen. Im Rahmen einer Anbieterbefragung adressiert Artikel II diese Herausforderungen. Denn mitunter kann durch die Anzahl der Kunden der automatisierten Texterstellung gezeigt werden, ob algorithmischer Journalismus bereits auf organisatorischer Ebene institutionalisiert ist. Dabei ist die Klassifizierung in Marktphasen auch für investitionsstrategische Entscheidungen im Nachrichtenmarkt von Bedeutung. Ist es notwendig als Medienhaus in diese Technologie zu investieren? Welche Produkte können angeboten werden? Nicht zuletzt kann sich auch die Beurteilung der Marktkonzentration je nach Marktphase unterscheiden und damit Entscheidungen zur Marktförderung oder Regulierung beeinflussen (vgl. Latzer et al. 2014). Märkte algorithmischer Selektion zeichnen sich zudem durch Merkmale der Qualität und Innovation aus, und weniger durch Preissetzung (vgl. Latzer et al. 2014: 13). Die Faktoren «Technisches

Know-how», «Qualität der Software-Plattformen und Algorithmen» wie auch die «Qualität der Daten» spielen im algorithmischen Journalismus eine zentrale Rolle und werden in Artikel II und III adressiert.

Nach Heuss (1965) durchlaufen Märkte im Laufe ihrer Entwicklung fünf Phasen. In der Experimentierphase wird das Produkt erfunden und zur Marktreife entwickelt. Mit der Markteinführung folgt die Expansionsphase, in der der Markt typischerweise exponentiell wächst. Diese beiden Phasen zeichnen sich im Allgemeinen durch hohe Konzentration und durch zeitbegrenzte Monopole von Innovatoren aus (vgl. Latzer et al. 2014: 14). Mit der Befragung der Software-Anbieter von algorithmischem Journalismus in Artikel II und der Erfassung von Faktoren wie dem Beschäftigungsgrad, dem Produktangebot, der Investitionssumme oder dem Kundenstamm können weitere Implikationen für die Marktstruktur, das Marktverhalten und das Marktergebnis gezogen werden. In der Ausreifungsphase ist das Wachstum noch auf hohem Niveau, jedoch gegenüber der Expansionsphase deutlich abgeschwächt. In der Stagnationsphase wird nur noch ein geringes bis überhaupt kein Wachstum mehr erreicht. In der Rückbildungsphase bildet sich der Markt zurück.

Die Ergebnisse der Anbieterbefragung werden zusätzlich durch die Ergebnisse der Journalistenbefragung in Artikel III ergänzt. Das heterogene Sample der Journalisten, bestehend aus Reportern, aber auch aus Journalisten in Managementpositionen, ermöglicht es, sie nicht nur in ihrer Rolle als professionelle Nachrichtenproduzenten zu befragen, sondern auch als Sprecher ihrer jeweiligen Organisation. Damit konnten weitere Indizien zur Diffusion von algorithmischem Journalismus auf Kundenseite für die Diskussion gewonnen werden. Gesondert wird hier auf die organisatorischen und technischen Imperative sowie auf die Diskussion der Forschungsergebnisse in Kapitel IV verwiesen.

Organisatorische und technologische Imperative

Organisatorische Imperative bestehen weitgehend aus Fragen zur Struktur und Organisation von Medienunternehmen und Redaktionen bei der Leistungserstellung (vgl. Weischenberg 1992: 275ff.). Professioneller Journalismus findet noch immer größtenteils in Medienorganisationen statt. Diese übernehmen zentrale Aufgaben und Leistungen des Journalismus (vgl. Weischenberg et al. 2006b: 346; Jarren 2008). Es ist daher zunächst einmal zu klären, welche technischen Strukturen für eine Integration der automatisierten Textproduktion in redaktionelle Abläufe notwendig oder welche bereits geschaffen sind (vgl. Hohlfeld 2010; Artikel II und III). Trotz der langen Tradition der Redaktionsforschung in der Kommunikationswissenschaft sind spezifische Fragestellungen zur Medien- und Redaktionsorganisation nicht explizit Bestandteil dieser Arbeit (vgl. Weischenberg 1992: 289). Es geht vielmehr darum, aufgrund der Neuartigkeit des Phänomens der algorithmischen Texterstellung mögliche Entwicklungen in diesem Bereich explorativ aufzuzeigen. Hier besteht weiterer Forschungsbedarf. Die redaktionelle bzw. organisatorische Struktur ist

nämlich besonders dann relevant, wenn es um die Diskussion ethischer und normativer Fragestellungen zur Transparenz, Kontrolle, Rechtmäßigkeit und Verantwortung im Umgang mit algorithmischem Journalismus geht (vgl. Artikel IV und V). Zudem stellt sich grundsätzlich die Frage, inwieweit auch journalistisch sozialisierte Akteure wie Software-Anbieter oder Programmierer journalistisch handeln können. Denn oftmals werden aufgrund von Spar- und Rationalisierungsmaßnahmen Teile journalistischer Arbeit einer Redaktion an externe Anbieter ausgelagert (vgl. Erdmann, Anhang A3; Blöbaum 2000; zur Übersicht und Diskussion des Akteursbegriffs in der Kommunikationswissenschaft Reine mann 2007: 50ff.; zur Kritik am systemtheoretischen Fokus und zu Vermittlungsversuchen zwischen Akteur- und Systemtheorie vgl. Hohlfeld 2003; Neuberger 2004, 2007; Quandt 2000; Schimank 2005; Wolf 2014). Dies führt ebenfalls zu Fragen der Relevanz, Aufgaben und Funktionen des redaktionellen Handelns im Spannungsfeld immer unabhängiger werdender Algorithmen und der Entwicklung im Machine Learning als Feld der Künstlichen-Intelligenz-Forschung. Die übergeordnete Frage, ob Redaktionen noch die zentralen Akteure im Gatekeeping oder Agenda Setting sind, ist nicht nur primär in Bezug auf algorithmischen Journalismus interessant, sondern vielmehr auch in Bezug auf Recommender-Systeme oder News-Aggregatoren auf unterschiedlichen Plattformen zu diskutieren (vgl. Just & Latzer 2016).

Implikationen für die mögliche organisatorische Ausrichtung von Redaktionen lassen sich einerseits über die Journalistenbefragung gewinnen (vgl. Artikel III). Auf qualitativer Ebene geben die Journalisten von Medien wie der BBC, The Sun oder Thomson Reuters Auskunft über Stand und Integration von automatisierter Texterstellung in ihren Unternehmen. Andererseits liefern auch die technische Analyse der Software und die Befragung der Software-Anbieter Hinweise auf die technischen Voraussetzungen der Redaktionsinfrastruktur (vgl. Artikel II). Über die wichtige Rolle der Technik in der journalistischen Aussagenproduktion besteht mittlerweile kein Zweifel (vgl. Weischenberg 1995). Das Wechselspiel zwischen Mensch und Technik, zwischen Redaktion und Algorithmus, ist immer dann Gegenstand der Forschung, wenn sich soziale oder organisationale Strukturen, Rollenbilder, Routinen oder Produktionsprozesse im Journalismus verändern (vgl. Mitchelstein & Boczkowski 2009; Powers 2012). Speziell für den algorithmischen Journalismus ergeben sich Chancen und Risiken bei der Auswahl, der Produktion und der Distribution der Inhalte. Die Analyse dieser Chancen und Gefahren und der Leistungsfähigkeit der Technik in der Nachrichtenproduktion wird aus technischer Perspektive in der Potentialanalyse und der Journalistenbefragung aufgegriffen (vgl. Artikel II und III). Sie bilden damit die Grundlage für die Strukturierung des Forschungsfeldes im Computational Journalism (Artikel I) sowie für die Diskussion der ethischen und ökonomischen Herausforderungen von algorithmischem Journalismus (Artikel III, IV und V).

3.3 Ebene der Medianaussagen

Wenngleich die Ebene der Medianaussagen (Funktionskontext) nicht explizit Teil der empirischen und theoretischen Analyse der Arbeit ist, wird der Einfluss von algorithmischem Journalismus auf sie dennoch in Abhängigkeit von und in Wechselwirkung mit der gesellschaftlichen, organisatorischen und professionellen Ebene sichtbar gemacht und in die Diskussion der Forschungsergebnisse in Kapitel IV einbezogen. Dies nicht zuletzt auch deshalb, weil algorithmischer Journalismus aufgrund seines spezifischen Leistungsvermögens explizit auf den Funktionskontext des Journalismus einwirkt und damit zeigt, welche Kommunikationsleistungen er bei der Beschreibung und Beobachtung der Gesellschaft erbringen kann. Die Dissertation analysiert dabei erstmals die Leistungen der Technik im Rahmen des Bezugssystems Journalismus in Artikel II (vgl. Weischenberg 1995: 110ff.).⁸ Die Funktion des algorithmischen Journalismus hat damit einen normativen Charakter (zum Problem der Abgrenzung von Funktion und Leistung siehe Godulla 2017: 51ff.; Neuberger 2008). Für die Beantwortung der übergeordneten Forschungsfrage wird die Ebene der Medianaussagen daher auch im Hinblick auf die Diskussion von Wechselwirkungen zwischen den Ebenen des Kontext-Modells in Kapitel IV relevant.

An dieser Stelle können nicht alle Facetten der Ebene der Medianaussagen dargestellt werden. Denn aufgrund der wachsenden wissenschaftlichen Erkenntnisse ist es unmöglich, alle Theorien und Befunde erschöpfend darzustellen. Die folgende Darstellung ist dabei an die Faktoren geknüpft, die bei der Diskussion in Kapitel IV aufgegriffen werden und sich auf Basis der empirischen und theoretischen Befunde der Arbeit herauskristallisiert haben. Wie bereits skizziert werden im «Funktionskontext des Journalismus» die Leistungen und Wirkungen des Journalismus beschrieben (vgl. Weischenberg 1992: 69). Faktoren, die im Zusammenhang mit den Medianaussagen stehen, sind u.a. journalistische Berichterstattungsmuster und Darstellungsformen (Genres), auf die die Artikel I und II verweisen. Des Weiteren ist der Faktor «Sprache und Verständlichkeit» zu nennen, bei dem es um die sprachliche Qualität der Medianaussagen geht. Damit einher gehen auch Funktionserwartungen der Rezipienten oder auch der Einfluss verschiedener Bezugsgruppen, wie beispielsweise der Journalisten oder der Software-Anbieter (vgl. Artikel I, II, III; Weischenberg 1995: 98f.). Erste Rezeptionsstudien zeigen, dass Leser kaum einen Unterschied zwischen manuell und automatisiert erstellten Texten erkennen können (vgl. Clerwall 2014; Graefe et al. 2016). Die Aussagenerstellung ist geprägt durch komplexe Handlungsabläufe,

⁸ Funktions- und Leistungsbeschreibungen des Journalismus können folgendermaßen voneinander abgegrenzt werden: «Leistungen in diesem Konzept sind [...] die Arbeiten, die gesellschaftliche Teilsysteme im Zuge funktionaler Differenzierung der Gesellschaft für die anderen Teilsysteme erbringen müssen. Die Leistung von Publizistik als System ist das Veröffentlichen von Themen gesellschaftlicher Kommunikation [...]. Funktionen dagegen sind nichts anderes als Lösungen eines zentralen Bezugsproblems der Gesamtgesellschaft (soziale Ebene 1) durch ein spezialisiertes Funktionssystem (soziale Ebene 2)» (Gehrke & Hohlfeld 1995: 281).

auf die unterschiedliche Akteure mit spezifischen Interessen und Rollenkonzeptionen Einfluss nehmen (vgl. Artikel II und III). Weitere Fragen zur Aussagenerstellung umfassen auch die Nachrichten- und Quellenauswahl bis hin zu ethischen Fragen der Objektivität, die besonders für den Journalismus aus normativer Perspektive bei der Konstruktion von Wirklichkeit relevant ist. Dabei findet die journalistische Wirklichkeitskonstruktion im Wechselspiel mit verschiedenen Akteuren statt. Fragen zur Verzerrung dieser Wirklichkeit stellen sich insbesondere auch durch neue Techniken wie den algorithmischen Journalismus. Artikel I, II und III thematisieren, wie sich der Einfluss von algorithmischem Journalismus insbesondere auch für den Bereich der Berichterstattungsmuster und Darstellungsformen im Computational Journalism auswirkt. Der Einfluss und das Wechselspiel von algorithmischem Journalismus mit der Ebene der Medienaussagen werden in Kapitel IV zudem in Bezug auf Sprache und Verständlichkeit sowie ethische Fragestellungen zur Nachrichtenauswahl diskutiert.

3.4 Ebene der Medienakteure

Auf professioneller Ebene (Medienakteure: Rollenkontext) sind journalistische Akteure über Arbeitsrollen an der Medienproduktion beteiligt (vgl. Weischenberg 1995: 373; Weischenberg et al. 2006b: 347). Algorithmischer Journalismus wirkt dabei auf unterschiedliche Faktoren dieser Ebene ein. Dabei stellt sich grundsätzlich die Frage, wie man dieses Wechselspiel analysierbar machen kann. Dies soll mit Verweis auf das Konzept des gradu-
alisierten Handelns möglich und in Kapitel IV aufgegriffen werden (vgl. Rammert & Schulz-Schaeffer 2002). Denn die gesellschaftliche Relevanz und der Beitrag von Technik zu sozialem Handeln verstärken sich im automatisierten Journalismus (vgl. Matsuzaki 2011: 301). In diesem Zusammenhang stellen sich auch Fragen nach der Substitution journalistischer Arbeit aus ökonomischer Perspektive. Insbesondere dann, wenn professionelles journalistisches Handeln automatisiert und dadurch verdrängt wird (vgl. Neuberger & Nuernbergk 2015; Weischenberg 1985: 187). Denn es ist gerade diese Beschreibungsleistung und -fähigkeit, die zentral dem Menschen zugeschrieben wird. Auch bei der Untersuchung von Produktionsprozessen und der Rolle professioneller Akteure im Journalismus bezieht man sich letztlich auf Menschen als handelnde Entitäten (vgl. Hohlfeld 2003: 120).⁹ Doch was Weischenberg und Hienzsch im Jahr 1994 für den Einsatz der Computertechnologie im Journalismus festgestellt haben, trifft auch auf die aktuelle Diskussion im algorithmischen Journalismus zu. Es verschwinden «nicht nur die Grenzen zwischen den Medien, sondern auch die Schnittstellen zwischen den Produktions- und Übermittlungs-

⁹ Als professionelle Akteure werden auch Redaktionen als Organisationen analysiert (vgl. Reinemann 2007).

techniken» (Weischenberg & Hienzsch 1994: 478). Im Folgenden werden die Problemdimensionen vorgestellt, die in direktem Bezug zu der Ebene der Medienakteure stehen und im Rahmen dieser Arbeit diskutiert werden.

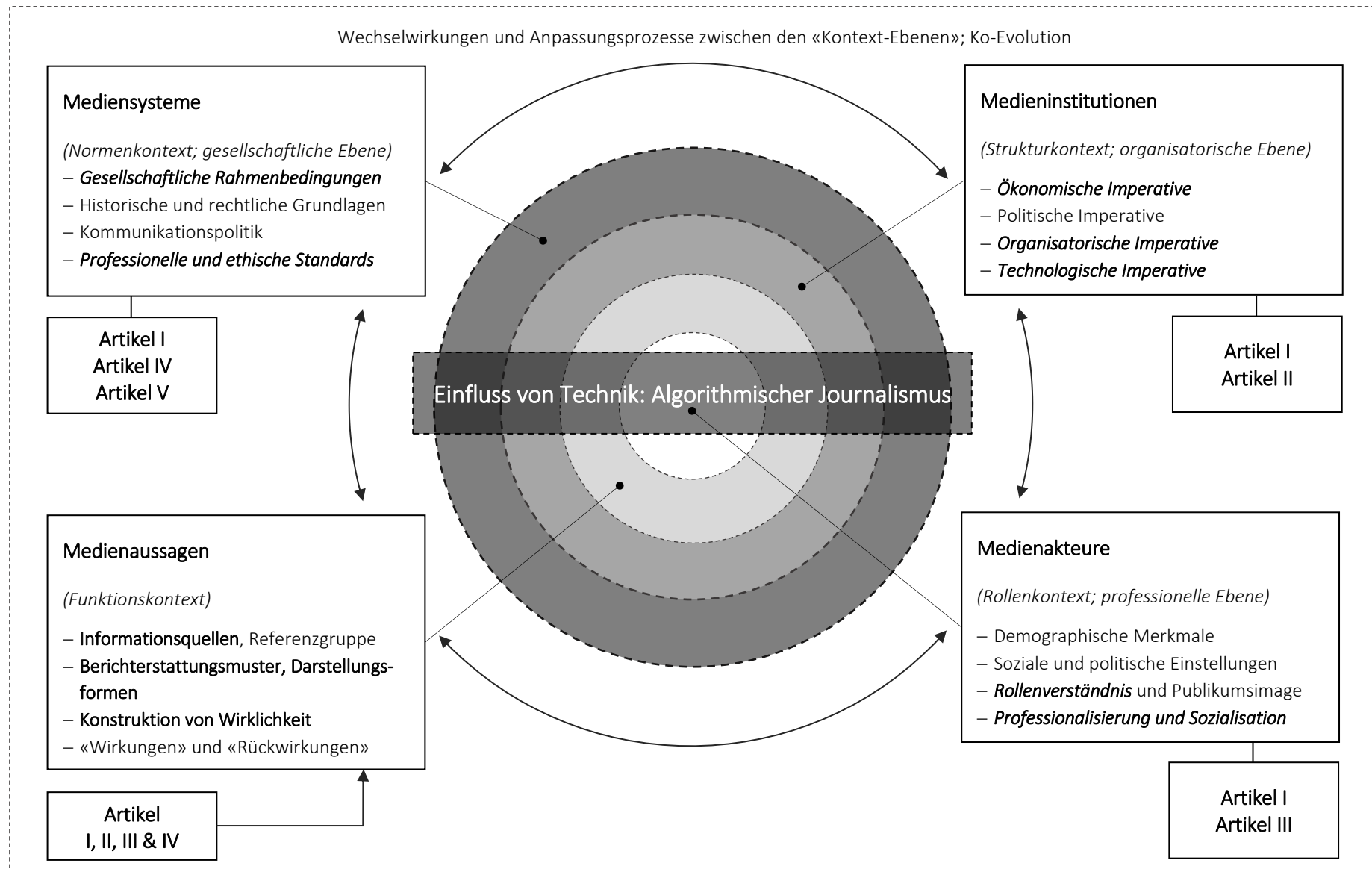
Auf der Ebene der Medienakteure diskutieren Artikel I, II und III insbesondere Fragen zur Rolle der Software-Anbieter, Programmierer und des Algorithmus im automatisierten Journalismus. Welche technischen Fähigkeiten sind relevant? Welche Arbeitsschritte sind nötig, um automatisiert Texte zu produzieren? Damit ergeben sich unweigerlich Fragen zum Status des Akteurs im Journalismus und übergeordnet die Frage, wie sich dadurch journalistische Tätigkeitsprofile und Anforderungen in Bezug auf den Beruf des Journalisten verändern können. Wie Journalisten algorithmischen Journalismus einschätzen, welche Chancen und Risiken sie sehen und wie sie ihre Aufgabe dabei definieren, hängt auch stark von ihrem Rollenselbstbild und -verständnis ab (vgl. Weischenberg 1995: 438). Rückschlüsse auf diese und die oben skizzierten Problemdimensionen werden im Workshop und in der Befragung der Journalisten in Artikel III deutlich. Übergeordnet werden auch die Ergebnisse der Potentialanalyse und der Strukturierung des Forschungsfeldes mit Verweis auf das Konzept der gradualisierten Handlungsträgerschaft sowie die Anbieterbefragung einbezogen (Artikel I, II und III). Denn der Einfluss von algorithmischem Journalismus auf die Ebene der Medienakteure kann nicht eindimensional diskutiert werden, sondern ausschließlich immer im Wechselspiel mit der gesellschaftlichen und organisatorischen Ebene.

Mit der Ebene der Medienakteure ist auch die gesellschaftliche Ebene verknüpft, und damit die ethische Diskussion um die Übernahme von Verantwortung und normative Qualität. Wie kann der Journalist die Leistungen und Aufgaben des Journalismus im Spannungsfeld der fortschreitenden Digitalisierung weiterhin erfüllen? Fragen zur Qualität der automatisierten Textproduktion hängen ebenfalls eng mit der technischen Kompetenz der Produzenten zusammen. Software-Anbieter und Journalisten haben damit direkten Einfluss auf zentrale Leistungen des Journalismus. Die Arbeit erkennt die zentrale Rolle dieser Akteure und adressiert die spezifischen Fragestellungen in Artikel I, II und III:

- a) *Wie lassen sich die Varianten des Computational Journalism voneinander abgrenzen?*
- b) *Hat Natural Language Generation (NLG) das technische Potential, traditionelle Funktionen des Journalismus in der Nachrichtenberichterstattung zu übernehmen?*
- c) *Wie schätzen Journalisten die technischen, ökonomischen, ethischen und gesellschaftlichen Chancen und Risiken von algorithmischem Journalismus aus professioneller Perspektive ein?*

Zusammenfassend visualisiert Abbildung 1 den Einfluss der automatisierten Textproduktion auf den Journalismus im Rahmen des Kontext-Modells nach Weischenberg (1992) und verknüpft die gesellschaftliche, organisatorische und professionelle Ebene mit den Untersuchungsdimensionen der Artikel I bis V.

Abbildung 1: Verortung und Einfluss der automatisierten Textproduktion im Journalismus in Anlehnung an Weischenberg (1992; et al. 2006), Esser (1998: 27) und Latzer (2013)



4. Zusammenfassung und Überblick

Zunächst wurde die automatisierte Textproduktion mit Verweis auf Latzer et al. (2014) als Anwendung algorithmischer Selektion und als Triebkraft im Medienwandel operationalisiert. Im Zuge dieser Entwicklung und im Spannungsfeld des Internets als Datenpool und als Ordnungs- und Regelsystem wurden verschiedene Herausforderungen für algorithmischen Journalismus auf gesellschaftlicher, organisatorischer und professioneller Ebene identifiziert. Um diese und die Wirkungsweise von algorithmischem Journalismus für eine kommunikationswissenschaftliche Analyse auf Mikro-, Meso- und Makroebene greifbar zu machen, wurde zunächst auf das Konzept von Technik als Akteur und Institution zurückgegriffen. Speziell das Konzept der gradualisierten Handlungsträgerschaft von Rammert und Schulz-Schaeffer (2002) eignet sich, um das Wechselspiel und die Leistungen von Journalist und Algorithmus in der Nachrichtenproduktion auf professioneller Ebene zu analysieren (vgl. Artikel I und II). Technik als Institution rückt ökonomische Fragen zum Markt der automatisierten Textproduktion in den Vordergrund. Mit der Übersicht und Einteilung der Anwendung in Marktphasen nach Heuss (1965) liefert die Dissertation dazu erste Einschätzungen. Auch organisatorische und ethische Fragen zu Arbeitsprozessen oder zur Kennzeichnung der automatisierten Inhalte können ein institutionelles Verständnis von Technik sichtbar machen (Artikel III, IV und V). Dieses konzeptionelle Verständnis zur Wirkung von Technik ermöglicht die Analyse unterschiedlicher Einflussfaktoren auf gesellschaftliche Prozesse. Denn das technische Potential von algorithmischem Journalismus regelt nicht nur die Leistungsfähigkeit, sondern hat auch Auswirkungen auf den Umgang mit ihm (vgl. Neuberger 2013: 103). Die Fragen zum Umgang mit der Technik wurden primär auf professioneller Ebene adressiert (Artikel III). Für die zukünftige Forschung sollte aber auch die Rolle des Publikums in die Analyse des algorithmischen Journalismus einbezogen werden (vgl. Hohlfeld 2016; Loosen 2016). Denn das Publikum entscheidet durch seine Leseauswahlentscheidungen letztlich, ob algorithmischer Journalismus auf organisatorischer Ebene diffundiert.

Die skizzierten Herausforderungen wurden im Anschluss mit den kommunikationswissenschaftlichen Fragestellungen der Dissertation verknüpft. Das Kontext-Modell von Weischenberg (1992) eignet sich dabei als heuristischer Analyserahmen, um den Einfluss der automatisierten Textproduktion auf journalistisches Handeln zu verorten. Damit war es möglich, die thematischen Schwerpunkte und die spezifischen Forschungsfragen der Artikel I bis V mit den Einflussfaktoren hinsichtlich journalistischen Handelns auf gesellschaftlicher, organisatorischer und professioneller Ebene des Journalismus zu verknüpfen. Es zeigt sich aber bereits bei der Zuordnung und den Ableitungen der Fragestellungen, dass die Herausforderungen nicht immer trennscharf auf eine Ebene bezogen werden können. Diese Schwierigkeit deutet bereits auf die Durchlässigkeit und das Wechselspiel der verschiedenen Faktoren auf den Ebenen hin und wird in Kapitel IV weiter ausgeführt.

Tabelle 1 liefert sowohl einen Überblick über die spezifischen Fragestellungen der Artikel I bis V als auch eine Zusammenfassung des spezifischen wissenschaftlichen Beitrags der Dissertation für die Analyse des Einflusses der automatisierten Textproduktion im Journalismus. Beides wird zusätzlich in Relation zur Analyseebene des Kontext-Modells gesetzt. Die Diskussion der Beiträge und die Beantwortung der übergeordneten Forschungsfrage dieser Arbeit folgen nach der Darstellung der methodischen Zugänge in Kapitel IV.

Tabelle 1: Übersicht über Forschungsfragen, den wissenschaftlichen Beitrag und Kontext-Ebenen als Bezugspunkte der Fragestellungen

Artikel	Forschungsfragen	Zentraler Beitrag	Bezug auf Ebene
Artikel I ₁₀	a) Wie lässt sich das Forschungsfeld zu Computational Journalism strukturieren? Was sind die zentralen Begriffe und Ausprägungen des Computational Journalism? b) Wie lassen sich die Varianten des Computational Journalism voneinander abgrenzen? c) Welche normativen Herausforderungen von Big Data und Algorithmen werden in der Nachrichtenproduktion auf den Ebenen der Datenbeschaffung, -auswertung und -präsentation relevant?	a) Überblick & Strukturierung des Forschungsfeldes für Lehre & Forschung; Definition & Abgrenzung zentraler Begriffe: Computational Journalism als Sammelbegriff – Differenzierung in CAR, Data Journalism & Algorithmic Journalism b) Systematisierungsvorschlag zur Abgrenzung der Varianten des Computational Journalism (gradualisierte Handlungsträgerschaft von Technik): Rolle des Journalisten, Rolle digitaler Daten & Rolle des Algorithmus; Übertragung auf E-V-A-Modell: Rolle & Aufgaben des Journalisten verändern sich mit steigendem Automatisierungsgrad & steigender Verfügbarkeit digitaler Daten c) Identifikation, Überblick & Diskussion normativer Herausforderungen: Transparenz, Objektivität, Rechtmäßigkeit & Kontrolle	Mediensysteme Medieninstitutionen Medienaussagen Medienakteure
Artikel II ₁₁	a) Hat Natural Language Generation (NLG) das technische Potential traditionelle Funktionen des Journalismus in der Nachrichtenberichterstattung zu übernehmen? b) Wer sind die zentralen Software-Anbieter von algorithmischem Journalismus (AJ)? c) In welcher Marktphase befindet sich algorithmischer Journalismus (AJ)?	a) Verortung im Forschungsprogramm «Algorithmische Selektion im Internet»; Technikpotentialanalyse von NLG als Basis von AJ (Literaturanalyse & Experteninterviews); Ableitung & Anwendung zentraler Funktionen & Leistungen des Journalismus auf NLG; Definition von AJ; AJ übernimmt Funktionen des faktenbasierten Informationsjournalismus, Limitierungen im Bereich der Kontextualisierung/Meinungsbildung b) Identifikation zentraler Software-Anbieter in DE, FR, GB, USA, CHN & Überblick über Gründungsjahr, Software-Produkte, Mitarbeiter, Kunden, Finanzierungsgrundlage c) Einordnung der Marktphase nach Heuss (1965) auf Basis strukturierter Experteninterviews mit Software-Anbietern: Expansionsphase, aber wenige Anbieter von AJ (Stand: 11/2015)	Medieninstitutionen

¹⁰ Dörr, K. 2016. Algorithmen, Big Data und ihre Rolle im Computational Journalism. In *Journalismusforschung. Stand und Perspektiven*, herausgegeben von Meier, K., & Neuberger, C., 245–264. Baden-Baden: Nomos.

¹¹ Dörr, K. 2016. Mapping the Field of Algorithmic Journalism. *Digital Journalism*, 4(6), 700–722.

Fortsetzung Tabelle 1

Artikel	Forschungsfragen	Zentraler Beitrag	Bezug auf Ebene
Artikel III ₁₂	Wie schätzen Journalisten die technischen, ökonomischen, ethischen und gesellschaftlichen Chancen und Risiken von algorithmischem Journalismus aus professioneller Perspektive ein?	<p>a) <i>Ansatz</i>: erstmalige Analyse der Einschätzungen zu Herausforderungen von AJ auf professioneller Ebene durch Methoden-Kombination (praktischer Workshop & strukturierte Experteninterviews) sowie im Kontext unterschiedlicher Rollenkonzeptionen & Fähigkeiten; Journalist als zentraler Akteur und Experte</p> <p>b) <i>Technik</i>: Identifikation zentraler Limitierungen bei Datenanalyse, Kontextualisierung/Meinung, Kreativität; aber: Möglichkeiten für Verwertbarkeit & gerätespezifische Adaption (z.B. mobil)</p> <p>c) <i>Ökonomie</i>: Verbreitung von AJ abhängig von Datenverfügbarkeit & möglichem Automatisierungspotential im organisationalen Kontext (z.B. Option für Nachrichtenagenturen), Rationalisierung repetitiver Arbeit</p> <p>d) <i>Ethik</i>: AJ als Teil des Computational Journalism & Add-on des traditionellen Journalismus: Tiefe, Breite, Genauigkeit, Geschwindigkeit & Unmittelbarkeit verfügbarer Informationen nehmen zu; Möglichkeiten, Bias und Ungenauigkeit zu reduzieren; Problem für Verantwortungsebene & Verifikation von Nachrichten</p>	Medienakteure
Artikel IV ₁₃	<p>a) Welche Anforderungen stellen die Digitalisierung und die Verfügbarkeit von digitalen Daten im Journalismus an professionsethisches Handeln?</p> <p>b) Inwieweit erfüllt algorithmischer Journalismus normative Qualitätsansprüche der Verantwortung und Transparenz?</p>	<p>a) Überblick & Identifikation normativer Qualitätsansprüche von AJ; Diskussion der normativen Kriterien «Verantwortung» & «Transparenz» aus journalismusethischer Perspektive</p> <p>b) AJ beeinflusst Ebene der Verantwortung vor, während und nach der Nachrichtenproduktion; Transparenz als normatives Gebot auch auf AJ zu übertragen</p>	Mediensysteme

12 Thurman, N., Dörr, K., & Kunert, J. 2017. When Reporters get Hands-On with Robo-Writing. Professionals Consider Automated Journalism's Capabilities and Consequences. *Digital Journalism* [online first]. doi: [10.1080/21670811.2017.1289819](https://doi.org/10.1080/21670811.2017.1289819).

13 Dörr, K., Köberer, N., & Haim, M. 2017. Normative Qualitätsansprüche an algorithmischen Journalismus. In *Gesellschaft ohne Diskurs? Digitaler Wandel und Journalismus aus medienethischer Perspektive. Kommunikations- und Medienethik, 5*, herausgegeben von Stapf, I., Prinzing, M., & Filipović, A., 121–131. Baden-Baden: Nomos.

Fortsetzung Tabelle 1

Artikel	Forschungsfragen	Zentraler Beitrag	Bezug auf Ebene
Artikel V ₁₄	a) Wo und wie lassen sich ethische Herausforderungen für den algorithmischen Journalismus theoretisch verorten? b) Welche ethischen Herausforderungen für den algorithmischen Journalismus lassen sich auf professioneller, organisatorischer und gesellschaftlicher Ebene identifizieren?	a) Identifikation, Beschreibung, Verortung & Verknüpfung ethischer Herausforderungen für den AJ im Spannungsfeld von Medien- von Cyberethik b) Entwicklung & theoretische Ableitung eines Analyserasters für die Identifikation ethischer Herausforderungen auf Mikro-, Meso- und Makroebene (Individuethik & Institutionenethik) in Anlehnung an Pürer (1992) & Weischenberg et al. (2006a,b); Analyse der ethischen Herausforderungen auf professioneller, organisatorischer und gesellschaftlicher Ebene aus individual- und sozialetischer Perspektive; Verknüpfung der Analyse mit der technischen Funktionsweise von AJ nach E-V-A-Modell; Fokus: Programmcode & Daten	Mediensysteme

¹⁴ Dörr, K., & Holnbuchner, K. 2016. Ethical Challenges of Algorithmic Journalism. *Digital Journalism* [online first]. doi: [10.1080/21670811.2016.1167612](https://doi.org/10.1080/21670811.2016.1167612).

III. Methodische Zugänge

In diesem Kapitel werden die methodischen Zugänge der Arbeit vorgestellt. Dabei erschweren die heterogene Definition von Journalismus, die vielfältigen theoretischen Zugänge und der Einfluss einer stetig sich weiterentwickelnden Technik sowohl den theoretischen wie auch den konkreten empirischen Zugang zum Untersuchungsgegenstand (vgl. Löffelholz & Rothenberger 2016; Scholl 1997). Doch in methodischer Hinsicht ist die Journalismusforschung seit jeher ein multiperspektivisches Forschungsfeld (vgl. Hanitzsch & Engesser 2014).

Der hier gewählte Forschungsansatz der Fallstudie eignet sich besonders, wenn explorativ komplexe, bisher wenig erforschte technische Phänomene wie der algorithmische Journalismus mit einem breiten Zugang und im Kontext seiner Einflüsse auf gesellschaftlicher, organisatorischer und professioneller Ebene des Journalismus operationalisiert und analysiert werden (vgl. Häder 2010: 350ff.). Diese Form der Mehrebenenanalyse eignet sich zur Modellierung dieser Einflussbeziehungen und hat – wie Hanitzsch und Engesser (2014: 150) zusammenfassen – vereinzelt bereits Einzug in die Journalismusforschung gehalten.

Die Arbeit verortet sich dabei in der Tradition der qualitativen Sozialforschung und ermöglicht mittels Methodentriangulation von Potentialanalyse, Befragung und Beobachtung eine differenzierte Betrachtung des Untersuchungsgegenstandes (vgl. Flick 2008; Lamnek 2010: 245ff.). Die Befragung und Beobachtung zählen dabei zu den zentralen Erhebungsmethoden von Fallstudien (vgl. Borchardt & Göthlich 2009; Lamnek 2010: 272ff.).

Die qualitativen Daten wurden in Artikel II und III eigenständig und originär erhoben und ermöglichen die explorative Systematisierung und Strukturierung des Forschungsfelds im Computational Journalism (Artikel I). Gleichzeitig liefern die Erhebungen den Startpunkt für die theoretische Diskussion aus normativer und ethischer Perspektive (Artikel IV und V). Dieses Vorgehen erlaubt es, den Forschungsgegenstand «algorithmischer Journalismus» multiperspektivisch zu betrachten und für zukünftige Fragestellungen zu öffnen. Die Dokumentation der methodischen Zugänge findet ausführlich in den Artikeln I bis V statt. In diesem Kapitel werden daher die einzelnen Ansätze miteinander verknüpft.

Zunächst wird das Instrument der Fallstudie vorgestellt. Im Anschluss daran werden die empirischen Einzelfallstudien, die Potentialanalyse, die Expertenbefragung von Software-Anbietern und Journalisten sowie die Beobachtung in den Artikeln II und III erläutert. Die kritische Betrachtung dieses methodischen Vorgehens rundet das Kapitel ab.

1. Die Fallstudie

Die Fallstudie wird in dieser Arbeit als übergeordneter Forschungsansatz, als «Metamethode», und nicht als eine spezifische Erhebungstechnik verstanden (von Rimscha & Sommer 2016: 373; Lamnek 2010: 299). Generell ist das Ziel von Fallstudien, ein «ganzheitliches und damit realistisches Bild der sozialen Welt zu zeichnen» (Lamnek 2010: 273). Einzelfallstudien eignen sich damit besonders für die explorative Analyse von algorithmischem Journalismus, indem sie Ursache-Wirkungs-Ketten aufdecken. Die Fallstudie versucht «Einblick in das Zusammenwirken einer Vielzahl von Faktoren [...] zu erhalten, wobei sie meist auf das Auffinden und das Herausarbeiten typischer Vorgänge gerichtet ist» (Fuchs et al. 1978: 181, zitiert nach Lamnek 2010: 274; Yin 2009). Die Deduktion der Fallstudienresultate auf ein größeres Ganzes ist möglich (vgl. von Rimscha & Sommer 2016). Dies ist auch Ziel der vorliegenden Arbeit. Sie beschreibt und skizziert dabei explorativ anhand von zwei Einzelfallstudien den Einfluss der automatisierten Textproduktion auf den Journalismus.

Mittels Methodentriangulation – also der Verwendung mehrerer Erhebungsmethoden und unterschiedlicher Untersuchungsobjekte – kann algorithmischer Journalismus aus unterschiedlichen Perspektiven betrachtet werden. So ist es möglich, seinen Einfluss auf gesellschaftlicher, organisatorischer und professioneller Ebene innerhalb des Journalismus sichtbar zu machen. Durch die Auswahl der Erhebungsmethoden «Potentialanalyse», «Befragung» und «Beobachtung» sowie den Fokus auf unterschiedliche Erhebungssituationen und Untersuchungsobjekte wird sowohl dem Postulat der Gegenstandsangemessenheit von Erhebungsmethoden als auch dem Postulat der Offenheit Rechnung getragen (vgl. Borchardt & Göthlich 2009: 44). Doch diese Entscheidung bei der Auswahl der Themenkomplexe «Technik» (Artikel I, II und III), «Ökonomie» (Artikel II und III) und «Normativität/Ethik» (Artikel III, IV und V) sowie bei der Auswahl der Untersuchungsobjekte «Technik», «Software-Anbieter» und «Journalisten» führt zwangsläufig auch zu einem Ausschluss anderer Ansätze. Offenen Fragen, insbesondere die vernachlässigte Beachtung des Publikums, werden daher gesondert in Kapitel V adressiert. Die Gültigkeit der gewonnenen Erkenntnisse wird aber – wie bei den Kooperationen in Artikel III, IV und V – erhöht, wenn mehrere Forscher im Forschungsprozess beteiligt werden. So können die gewonnenen Ergebnisse untereinander und vor dem Hintergrund unterschiedlicher theoretischer Ansätze diskutiert werden (vgl. Yin 2009). Dies war insbesondere bei der Diskussion der normativen und ethischen Herausforderungen hilfreich in Artikel IV und V.

1.1 Potentialanalyse

Um die technischen Eigenschaften der automatisierten Texterstellung in Artikel II zu evaluieren, nutzt die Arbeit das methodische Instrument der Technikanalyse. Diese Methode

der Technikfolgenabschätzung wird schon seit den 1960er und 1970er Jahren für die Evaluation von Technik und Innovationen genutzt (vgl. Decker et al. 2012; Grunwald 2010; Häußling 2014). Der Einsatz dieses Instruments ist in der Journalismusforschung jedoch noch nicht weit verbreitet (siehe Neuberger 2001; Wolf 2014). Technik wird hier jedoch nicht deterministisch, sondern im Sinne des koevolutionären Ansatzes analysiert. Dieses Verständnis ermöglicht dabei auch einen Rückgriff auf die skizzierte theoretische Operationalisierung von Technik als Akteur und das Konzept der gradualisierten Handlungsträgerschaft (vgl. Rammert & Schulz-Schaeffer 2002). Damit konnte Artikel I die unterschiedlichen Varianten des Computational Journalism mit Verweis auf die Rolle der Journalisten, die Rolle von Algorithmen und die Rolle der Daten in der Nachrichtenproduktion differenzieren und das Forschungsfeld zunächst strukturieren. Dieses Vorgehen und die Differenzierung des techno-sozialen Wechselspiels in verschiedene Handlungsebenen ermöglichen die Analyse der Rolle der Technik auch für ethische Anschlussfragen zu öffnen (siehe Artikel IV und V).

Die Potentialanalyse von algorithmischem Journalismus in Artikel II baut dabei sowohl auf einer ersten literaturgeleiteten Vorstudie als auch auf einer Expertenbefragung auf (siehe u.a. Anhang A). Daraus wurden Themen und Fragenkomplexe für die Anbieterbefragung abgeleitet. Auch die Ergebnisse der qualitativen Befragung und Beobachtung in Artikel III – Journalisten arbeiten erstmals mit einer marktreifen Softwarelösung – ergänzen diese Ergebnisse aus einer weiteren Perspektive. Dieser mehrdimensionale Blickwinkel führt dabei nicht zu einer einseitigen Interpretation der Chancen und Risiken von algorithmischem Journalismus und erlaubt gleichzeitig eine wertfreie Identifikation und Diskussion der normativen und ethischen Herausforderungen in Artikel IV und V.

1.2 Experteninterview

Der zentrale forschungsmethodische Ansatz der Dissertation ist das halbstandardisierte Experteninterview. Es hat sich in einer Vielzahl von Studien im Journalismus als Erhebungsinstrument etabliert (vgl. u.a. Meyen 2009). Diese Methode wird in der Arbeit als Grundlage für die Einschätzung des technischen und ökonomischen Potentials von algorithmischem Journalismus (Artikel II) sowie für die Befragung von Journalisten zu technischen, ökonomischen und ethischen Aspekten der automatisierten Texterstellung genutzt (Artikel III). Zugleich dient das Experteninterview auch der Kontextualisierung und Validierung der Ergebnisse der Potentialanalyse und befasst sich mit Wissenschaftlern und Software-Anbietern sowie deren theoretischem Wissen über die automatisierte Texterstellung (vgl. Hoffmann 2005: 269f.). Das Erhebungsinstrument wird in der Arbeit dabei sowohl explorativ als auch systematisierend eingesetzt. Es ermöglicht eine erste Orientierung im Forschungsfeld wie auch eine Erhebung des Sachwissens der Experten (vgl. Bogner et al. 2014). Wenngleich Experteninterviews oft nur am Rande und als spezifische Form von

qualitativen Interviews beschrieben werden (vgl. Blöbaum et al. 2016: 175ff.), hat das Experteninterview aufgrund seines Erkenntnisinteresses einen besonderen Status: Es «hat das Potential, die ‹Maximen, Regeln und Logiken› (Meuser & Nagel 2009: 472) des Handelns in Organisationen und gesellschaftlichen Teilbereichen zu identifizieren und somit die Funktionsweise von Gesellschaft zu analysieren» (Blöbaum et al. 2016: 176). Denn in dieser Arbeit stehen die Erfahrungen und Interpretationen der Experten zu algorithmischem Journalismus im Zentrum des Erkenntnisinteresses. Experten sind allgemein alle Akteure in Organisationen bzw. Rollenträger in gesellschaftlichen Teilbereichen mit Zugang zu Informationen über Personengruppen, Organisationsabläufe und Entscheidungsprozesse (vgl. Blöbaum et al. 2016; Bogner et al. 2014). Der Experte wird hier in Abgrenzung zum qualitativen Leitfadeninterview und in Anlehnung an Meuser & Nagel (2009) und Blöbaum et al. (2016) als Funktionsträger und nicht als Privatperson operationalisiert. «Der Experte steht für eine Problemperspektive, die typisch ist für den institutionellen Kontext, in dem er sein Wissen erworben hat und in dem er handelt» (Meuser & Nagel 2009: 469). Dieser Definition wird mit der Auswahl der Interviewpartner in der Arbeit Rechnung getragen. Wissenschaftler, Software-Anbieter (Artikel II) und Journalisten (Artikel III) werden als Experten in ihrem spezifischen Rollenkontext befragt. Die Interviews wurden dabei in deutscher und englischer Sprache Face to Face und via Skype in Artikel II sowie Face to Face in Artikel III mit einem halbstandardisierten Leitfaden geführt. Dies war insofern notwendig, als die Interviews in Artikel III von mehreren Forschern durchgeführt wurden und so eine Vergleichbarkeit der Interviewergebnisse gewährleistet werden konnte. Die Interviews wurden elektronisch aufgezeichnet und im Anschluss transkribiert. Sie bilden damit die Grundlage für die Interpretation der Ergebnisse.

1.3 Beobachtung

Die Arbeit nutzt mit der Beobachtung ein weiteres Instrument der Fallanalyse. Es wurde eine technisch vermittelte wie auch unvermittelte, teilstandardisierte, nichtteilnehmende qualitative Beobachtung durchgeführt (vgl. Artikel III; Borchardt & Göthlich 2009). Allgemein ist die Beobachtung «das systematische Erfassen, Festhalten und Deuten sinnlich wahrnehmbaren Verhaltens zum Zeitpunkt seines Geschehens» (Atteslander 2006: 67). Die Beobachtung wird dabei nicht als einzelne Erhebungsmethode genutzt, sondern in Kombination mit dem Instrument des qualitativen Experteninterviews. Im Rahmen eines Workshops hatten Journalisten ausgewählter englischer Medien die Möglichkeit, erstmals mit der Software eines führenden Anbieters zu arbeiten. Nach einer kurzen Einführung in das Software-System durch den Forschungsleiter erhielten die Teilnehmer die Aufgabe, eine Syntax (Code) für einen automatisierten Text aus den Bereichen Sport, Kriminalität und Finanzen zu schreiben. Der gesamte Workshop, der unter Laborbedingungen an der London City University stattfand, wurde durch eine verspiegelte Wand beobachtet. Die

Beobachter wurden dabei jenem Journalisten zugeteilt, mit dem im Anschluss auch das Interview geführt wurde. Die Beobachtung der Körpersprache durch den Spiegel sowie die elektronische Übertragung des Bildschirms und der Audiosignale machten es möglich, die spezifischen technischen Fähigkeiten der Journalisten einzuschätzen. Kombiniert mit den Interviews half die Beobachtung bei der Interpretation und Kontextualisierung der Aussagen zur Selbsteinschätzung der technischen Fähigkeiten. Das gesamte Potential der generierten Daten der Beobachtung (Audio und Bildschirmaufzeichnung) konnte bisher jedoch noch nicht vertieft für weitere Forschungsfragen, z.B. zu Skills und Usability der Software, genutzt werden. Der Mehrwert der Beobachtung ist damit eher in der Kontextualisierung und Interpretation der Forschungsergebnisse zu sehen als in der Verwendung als eigenständige Erhebungsmethode.

2. Zusammenfassung und Diskussion

Das Instrument der Fallstudie ist für eine erste explorative Analyse von algorithmischem Journalismus geeignet. Die Methodenkombination von Technikanalyse, Experteninterview und Beobachtung trägt zur Objektivität der Untersuchungsergebnisse auf organisatorischer und professioneller Ebene bei und hilft das Forschungsfeld für die theoretische Diskussion von Ethik und normativer Qualität auf gesellschaftlicher Ebene zugänglich zu machen. Die empirische Anknüpfung an die organisatorische (Artikel II) und professionelle Ebene (Artikel III) erlaubt es, den Einfluss von algorithmischem Journalismus auch für die Ebene der Medienaussagen mitzudiskutieren. Das zeigt sich besonders für die Diskussion der Leistungen von algorithmischem Journalismus in Artikel IV sowie für mögliche Muster in der Berichterstattung und bei den Darstellungsformen.

Generell ist es eine Schwäche der Expertenbefragung von Journalisten und Software-Anbietern, dass nur die Faktoren sichtbar gemacht werden können, die von den Expertengruppen auch als solche wahrgenommen werden (vgl. Hanitzsch 2009). Die Neuartigkeit der Technik, der Rollenkontext, technisches Wissen und Kompetenzen, ökonomische Interessen oder auch die grundlegende Sympathie zu den Interviewern beeinflussen die Aussagekraft der Ergebnisse. Die Ergebnisse der Journalistenbefragung in Artikel III müssen zudem im Kontext der Erhebung diskutiert werden. Obwohl das Sample sich aus Journalisten unterschiedlichen Alters sowie variierender technischer Fähigkeiten und professioneller Erfahrung zusammensetzt, sind die Meinungen, speziell zum technischen Potential der automatisierten Texterstellung, durch ihre Erfahrung mit der Software eines führenden Anbieters im Rahmen des Workshops beeinflusst. Rollenkontexte, technisches Wissen und die begrenzte Zeit, die sie mit der Software arbeiten konnten, beeinflussen damit die Aussagekraft der Ergebnisse. Dennoch war es einigen der Journalisten möglich, das Potential der Software in vollem Umfang auszutesten und mögliche Einschränkungen für eine

breite journalistische Nutzung festzustellen. Darüber hinaus versuchen die bewusste Auswahl der Journalisten und die Variation bei den Merkmalen «Position/Arbeitsbereich», «technische Fähigkeit» und «Medienorganisation» sowie die vorausgegangene Schulung durch den Untersuchungsleiter dieses Problem zu relativieren. Zudem wurden die Journalisten auch nicht nur als Experten der Nachrichtenproduktion, sondern auch als Vertreter ihrer Organisation befragt (Managementposition). Damit war es möglich, speziell zum Potential des algorithmischen Journalismus und zu seinem Einsatz gültige Aussagen abzuleiten.

Das Potential der Beobachtung als eigenständige Erhebungsmethode wurde noch nicht ausreichend genutzt und bietet die Möglichkeit für weitere Forschung. Dies trifft insbesondere für Fragen zur Usability der Texterstellungsoftware zu, da die Arbeit der Journalisten mit der Software vollständig elektronisch aufgezeichnet wurde. Hier könnte auch ein Vergleich unterschiedlicher Software-Systeme hilfreich sein.

Dennoch: Der Mehrwert dieser explorativen Herangehensweise der Arbeit liegt darin, den Einfluss von algorithmischem Journalismus zunächst einmal greifbar und für weitere Forschung zugänglich zu machen. Dieses Vorgehen ergibt sich aus dem Status seiner Diffusion im Journalismus und dem eingeschränkten wissenschaftlichen Forschungsstand zu diesem Phänomen.

IV. Diskussion der Forschungsergebnisse

Mit Verweis auf die übergeordnete Forschungsfrage dieser Arbeit werden in diesem Kapitel die zentralen Forschungsergebnisse der Artikel I bis V diskutiert. Es wird gezeigt, welchen Einfluss algorithmischer Journalismus auf gesellschaftlicher, organisatorischer und professioneller Ebene des Journalismus hat und welche Interdependenzen sich als Konsequenz zwischen diesen Ebenen identifizieren lassen. Um dies zu tun, orientiert sich die folgende Diskussion entlang der Ebenen des Kontext-Modells nach Weischenberg (1992) sowie an der techniksoziologischen Operationalisierung von Technik als Akteur und Institution. Die Ebene der Medienaussagen wird dabei als Resultat des Zusammenspiels von gesellschaftlicher, organisatorischer und professioneller Ebene auch in die Analyse einbezogen.

Welchen Einfluss hat algorithmischer Journalismus auf gesellschaftlicher, organisatorischer und professioneller Ebene des Journalismus und welche Interdependenzen lassen sich als Konsequenz zwischen diesen Ebenen identifizieren?

1. Einfluss auf die Ebene der Mediensysteme

Um die Wirkung von algorithmischem Journalismus auf die Ebene der Mediensysteme sichtbar zu machen, werden insbesondere die Ergebnisse der Potentialanalyse von Artikel II und die Systematisierung in Artikel I herangezogen. Denn aus dem technischen Potential der automatisierten Texterstellung lassen sich die skizzierten Fragestellungen sowohl ableiten als auch beantworten (vgl. Tabelle 1). Die Ebene der Mediensysteme liefert dabei allgemeine und spezifische Werte und Normen, die besonders für den normativen und ethischen Diskurs sowie die Qualität von Medienprodukten in Artikel I, III, IV und V eine Rolle spielen. Sowohl der Ansatz von Technik als Akteur (Potentialanalyse) als auch der Ansatz von Technik als Institution eignen sich für die ethische Analyse von algorithmischem Journalismus. Mit der Potentialanalyse von algorithmischem Journalismus lassen sich ethische Problemdimensionen zur Wirkung von Technik auf der Mikro-, Meso-, und Makroebene identifizieren.

Zunächst identifiziert Artikel IV Qualitätsmaßstäbe, die im journalistischen Handlungsfeld übergeordnet Geltung beanspruchen können. Ethische Kriterien wie die Achtung vor der Wahrheit, die Wahrung der Menschenwürde und die wahrhaftige Unterrichtung der Öffentlichkeit zählen dabei zu den Grundnormen des Journalismus (vgl. Wunden 2003: 175ff.; Köberer 2014). Diese Kriterien sind eine wichtige Voraussetzung für die freie Meinungsbildung sowie die Partizipation der Öffentlichkeit und damit eine Grundbedingung für die Umsetzung demokratischer Prinzipien durch den Journalismus (vgl. Kiefer 2010). Als Prinzipien journalistischer Qualitätssicherung lassen sich mit Bezug auf die

Transparenz auch die von Pöttker (1999) formulierten Trennungsgrundsätze nennen. Diese Normen zur Verhaltensorientierung finden sich auch in den «Publizistischen Grundsätzen», z.B. im Pressekodex des Deutschen Presserats (vgl. Weischenberg 2001: 279ff.).

Vor dem Hintergrund einer zunehmenden Ausdifferenzierung des journalistischen Handlungsfeldes und der Ergänzung durch algorithmischen Journalismus stellt sich die Frage, wem Verantwortung zugeschrieben wird und wer sie wie tragen soll. Artikel IV und V greifen dabei mit Verweis auf die Medienethik auf einen Verantwortungsbegriff zurück, bei dem verschiedene Verantwortungsebenen berücksichtigt werden. Die Artikel differenzieren daher das Mediensystem in verschiedene Handlungs- und Verantwortungsbereiche, u.a. von Medienmachern, Besitzern und Betreibern von Massenmedien sowie den Mediennutzern. Mit dieser Verteilung und Zuschreibung von Verantwortung aus individualethischer sowie aus sozialetischer Perspektive können einzelne Medienakteure sowie Akteure auf organisatorischer und gesellschaftlicher Ebene in die Analyse einbezogen werden. Zudem verortet Artikel I diese Herausforderungen erstmals im Rahmen der Medienethik und Cyberethik und hilft damit bei der Orientierung im Forschungsfeld.

Diese Herausforderungen im Umgang mit der automatisierten Textproduktion lassen sich zudem analog zum E-V-A-Modell auf der Ebene der Datenbeschaffung (Eingabe), der algorithmisch gesteuerten Datenauswertung im Programmcode (Verarbeitung) und der Präsentation der Inhalte (Ausgabe) feststellen. Dabei verorten Artikel I, Artikel IV und V die normativen und ethischen Herausforderungen von algorithmischem Journalismus im Rahmen der journalistischen Ethik und strukturieren die spezifischen Problemdimensionen auf gesellschaftlicher, organisatorischer und professioneller Ebene. Folgt man diesem groben Raster, zeigen sich Interdependenzen mit der organisatorischen und professionellen Ebene, aber auch mit der Ebene der Medienaussagen. Denn algorithmischer Journalismus beeinflusst die Anforderungen an die Akteure der professionellen und organisatorischen Ebene und ordnet die Verantwortungsebenen vor, während und nach der Nachrichtenproduktion neu (vgl. Artikel V). Der Journalist als professioneller Akteur ist dabei eingebettet in seinen organisationalen Kontext der Redaktion und Medienorganisation (vgl. Artikel III). Entscheidungen bei der Auswahl der Daten als Grundlage der automatisierten Textproduktion werden in diesem Kontext getroffen. Der Journalist ist damit nicht mehr die alleinige Instanz in diesem Auswahlprozess. Hierbei sind redaktionelle Entscheidungen bei der Themensetzung genauso einzubeziehen wie ökonomische Zwänge der Medienorganisation. Denn es eignet sich nicht jedes Thema für algorithmischen Journalismus, auch weil oftmals eine breite Datengrundlage fehlt.

Auf organisatorischer Ebene haben neben der Redaktion als professionellem Akteur auch vermehrt Software-Anbieter (Programmierer) Einfluss auf die inhaltliche Gestaltung

der Medienprodukte im algorithmischen Journalismus. Denn die Software ist der Ermöglichungsrahmen, die Enabling Technology. Je nachdem wie gut und wie vielfältig der programmierte Algorithmus ist, desto besser kann auch – die technischen Fähigkeiten der Journalisten und eine gute Datenlage vorausgesetzt – das finale Nachrichtenprodukt sein. Während die Anbieter dieser Software die Texte ihrer journalistischen Kunden noch zu Beginn der technischen Entwicklung vermehrt inhouse nach spezifischen Kundenwünschen produziert haben, lässt sich allmählich ein strategischer Wechsel feststellen. Oftmals wird nur noch der Zugang zur Software-Plattform bereitgestellt (SaaS) (vgl. Artikel III). Der Journalist kann nun selbstständig über ein Interface die Syntax für die automatisierten Texte erstellen. Auch die Datenbeschaffung liegt mehrheitlich in der Verantwortung der journalistischen Kunden. Mit dieser Entscheidung wird die Verantwortung im Produktionsprozess wieder auf die Medienmacher und Medienbesitzer der professionellen und organisatorischen Ebene verteilt. Ob die Kriterien der normativen Qualität wie Objektivität, Rechtmäßigkeit oder Transparenz befolgt werden, hängt nun von diesen Akteuren ab. Dennoch ergeben sich speziell im Hinblick auf das Gebot der Transparenz als zentrales Kriterium normativer Fragestellungen ungeklärte Fragen (u.a. Jungnickel 2011). Während die Entscheidung, ob Inhalte als automatisiert gekennzeichnet werden oder welche Daten als Grundlage der Textgenerierung dienen, in der Verantwortung der professionellen Akteure liegt, ist die Frage der vollständigen Transparenz des Programmcodes noch ungeklärt. Denn dieser unterliegt dem Geschäftsgeheimnis der Anbieter. Wie genau der Algorithmus nun den Daten eine bestimmte Relevanz zuordnet und auf welcher Basis er im Rahmen der Syntax bestimmte Inhalte präferiert, ist damit noch immer ungeklärt. Diese Ebene wird im Rahmen der Entwicklung im Bereich der künstlichen Intelligenz immer wichtiger. Denn gerade in der Phase der Verarbeitung zeigt sich einerseits die Qualität des algorithmischen Journalismus für die Nachrichtenproduktion und andererseits ergeben sich auch potentiell neue Fragestellungen bei der Diskussion von Verantwortung. Wie Artikel IV argumentiert, sollten aus medienethischer Perspektive automatisierte Inhalte im Sinne des Trennungsgrundsatzes – Unterscheidung zwischen redaktionellen Inhalten und Werbung – auch als solche gekennzeichnet werden. Dies ist eine Einschätzung, die auch von den meisten Journalisten in Artikel III geteilt wird. Sie identifizieren Transparenz als zentrales normatives Kriterium. Die Erkennbarkeit der Autorenschaft eines Beitrags ist dabei wichtig für den Anspruch der Transparenz als Kriterium journalistischer Qualität (vgl. Montal & Reich 2016). Dennoch bleibt die Frage, was diesem Anspruch genügt: die Offenlegung der Automatisierung, die Offenlegung der Datenquellen oder gar die Offenlegung des Programmcodes? Auch die Journalisten in Artikel III waren nicht einheitlicher Meinung und thematisierten die ethischen Herausforderungen primär im Rahmen ihrer Routinen in der Nachrichtenproduktion. Ökonomische Kriterien wie Umsatz und Gewinn spielen aus Sicht der Befragten in diesem Zusammenhang keine wesentliche Rolle. Ihrer

Einschätzung, dass algorithmischer Journalismus Bias und Ungenauigkeit in der Berichterstattung reduzieren kann, ist entgegenzuhalten, dass dieses Potential stark an die technischen Fähigkeiten bei der Programmierung, die Auswahl und Gestalt der Datenlage sowie die Qualität der Software rückgebunden ist.

Wie und ob den Kriterien der normativen Qualität in der Nachrichtenproduktion gefolgt wird, beeinflusst damit auch die Ebene der Medienaussagen und insofern die Leistungen und Wirkungen von algorithmischem Journalismus bei der Beobachtung und Beschreibung der Gesellschaft. Hier besteht weiterer Forschungsbedarf, da Medienorganisationen aktuell auch aufgrund der Neuartigkeit des Phänomens sich in einer Adaptionsphase befinden.

Bei normativen Fragen der Verantwortung gilt es insbesondere auch zu klären, ob Algorithmen als Träger von Verantwortung gelten können. Dies ist insofern relevant, als ethische Fragestellungen immer subjektgebunden sind und rationales Handeln bisher nur dem Menschen zugeschrieben wurde (vgl. Artikel IV). Artikel II und der Rückgriff auf das Konzept der gradualisierten Handlungsträgerschaft ermöglichen eine klare Antwort auf diese Frage. Dies bezieht sich jedoch nur auf den aktuellen Stand der Entwicklung der Technik und kann sich mit dem technischen Fortschritt wieder ändern: Algorithmen besitzen zwar gewisse Freiheitsgrade bei der Texterstellung, können aber nicht intentional handeln und sind in ihrem Gestaltungsspielraum an die Programmierung und die Auswahl der Daten gebunden (vgl. Artikel I und II; siehe auch Abschnitt 4: *Einfluss auf die Ebene der Medienakteure*). Die Frage der Verantwortung liegt damit auf organisatorischer und professioneller Ebene bei den Medienmachern und Medienbetreibern bzw. -besitzern, die die zentralen Leistungen des Journalismus bei der Beobachtung und Beschreibung der Gesellschaft garantieren müssen (vgl. z.B. § 55 Abs. 2 RStV¹⁵). Algorithmischer Journalismus setzt die Akteure damit aber auch unter Druck, da, wie Artikel IV und V argumentieren, Handlungsorientierungen und -anweisungen für den Umgang mit und die Produktion von journalistischen Inhalten auf organisatorischer wie individueller Ebene kontinuierlich angepasst werden müssen. Konkrete ethische Richtlinien sind – auch aufgrund der Neuheit der Technik – noch Mangelware und sind oftmals in organisationalen Richtlinien eingebettet (vgl. Anhang A3, A4).

Ob und wie die Kriterien normativer Qualität und Verantwortung auf der Ebene des Mediensystems erfüllt werden, hängt damit nicht zuletzt auch von der Beachtung dieser Grundsätze auf der Ebene der Medieninstitutionen und der Medienakteure ab. Die zentralen Beiträge von Artikel III, IV und V ermöglichen eine Verortung und erste explorative

¹⁵ § 55 Abs. 2 RStV: «[...] Anbieter von Telemedien mit journalistisch-redaktionell gestalteten Angeboten, in denen insbesondere vollständig oder teilweise Inhalte periodischer Druckerzeugnisse in Text oder Bild wiedergegeben werden, haben zusätzlich zu den Angaben nach den §§ 5 und 6 des Telemediengesetzes einen Verantwortlichen mit Angabe des Namens und der Anschrift zu benennen. [...]».

Analyse der ethischen Herausforderungen unter Berücksichtigung individual- und sozial-ethischer Ansätze sowie eines differenzierten Verantwortungsbegriffs. Mit «reverse engineering», also dem «process of extracting the knowledge or design blueprints from anything man-made» (Eilam 2005: 3; Diakopoulos 2015: 404), konnten die Potentialanalyse und die Strukturierung der Funktionsweise mit dem E-V-A-Modell in Artikel II maßgeblich zur Identifikation weiterer ethischer Problemdimensionen beitragen (siehe *Technik als Akteur*). Die Herausforderungen zeigen sich dabei primär auf der Ebene Daten und des Programmcodes. In diesem Sinne liefern Artikel III, IV und V mit der Verortung und theoretischen Diskussion den Startschuss für weitere Forschung.

2. Einfluss auf die Ebene der Medieninstitutionen

Der regelnde und ordnende Charakter von algorithmischem Journalismus wird mit Verweis auf das Konzept von Technik als Institution besonders auf der Ebene der Medieninstitutionen sichtbar. Wie bereits erwähnt, ist eine trennscharfe Verortung der spezifischen Fragestellungen der Artikel I bis V nicht immer möglich. Dennoch sollen mit Verweis auf die ökonomischen und technischen Imperative auf der Ebene der Medieninstitutionen der zentrale Beitrag und der Einfluss auf das Kontext-Modell diskutiert werden. Zunächst werden die technischen und organisatorischen Implikationen der Ergebnisse vorgestellt. Danach folgt die Diskussion der ökonomischen Imperative.

Wie die Forschungsergebnisse in Artikel II zeigen, übernimmt die automatisierte Textproduktion im Wechselspiel mit den Medienakteuren institutionelle Aufgaben des Journalismus bei der Beobachtung und Beschreibung der Gesellschaft (siehe insbesondere Abschnitt 3: *Einfluss auf die Ebene der Medienaussagen*). Diese Beobachtung und Beschreibung kann durchaus aktuell sein, falls die Datenbanken laufend aktualisiert werden können und der Programmcode entsprechend angepasst ist. Dies zeigt u.a. auch das Beispiel des Quakebots der LA Times, der über Erdbeben berichten konnte, weil er live auf die Datenbank des U.S. Geological Survey zugreift. Die Ergebnisse der Anbieterbefragung in Artikel III mit dem Hinweis auf den Kundenstamm lassen vermuten, dass diese Leistungen in den organisationalen Rahmen der Redaktion bzw. Medienorganisation eingebunden sind und in diesem Rahmen erbracht werden. Diese Integration befindet sich allerdings noch in einer frühen Phase und ist nicht wie bei anderen Anwendungen des Computational Journalism weit fortgeschritten (Artikel I).

Die redaktionelle Entscheidung, ob algorithmischer Journalismus genutzt wird, beeinflusst damit die gesellschaftliche Ebene sowie den Umgang und die Ausgestaltung der Medienprodukte. Artikel II und III zeigen, dass mit der Integration der Technik in redaktionelle Strukturen und Arbeitsabläufe aufgrund ihrer technischen Disposition nur bestimmte Berichterstattungsmuster und spezifische Darstellungsformen bedient werden

können. Diese sind größtenteils Aufgaben und Leistungen des Informationsjournalismus. Technische Imperative beeinflussen damit auch die Ebene der Medienaussagen (siehe Abschnitt 3: *Einfluss auf die Ebene der Medienaussagen*). Der Zugang zur Technik ist für Redaktionen aber unproblematisch und findet über die Cloud im Internet statt (Artikel II und III). Der Workshop in Artikel III zeigt auch, dass es zu Funktionsproblemen kommen kann, wenn Normen und Standards der Technik nicht eingehalten werden. Texte können sowohl sprachlich als auch inhaltlich unkorrekt sein. Dabei spielen technische Fähigkeiten der Journalisten eine entscheidende Rolle.

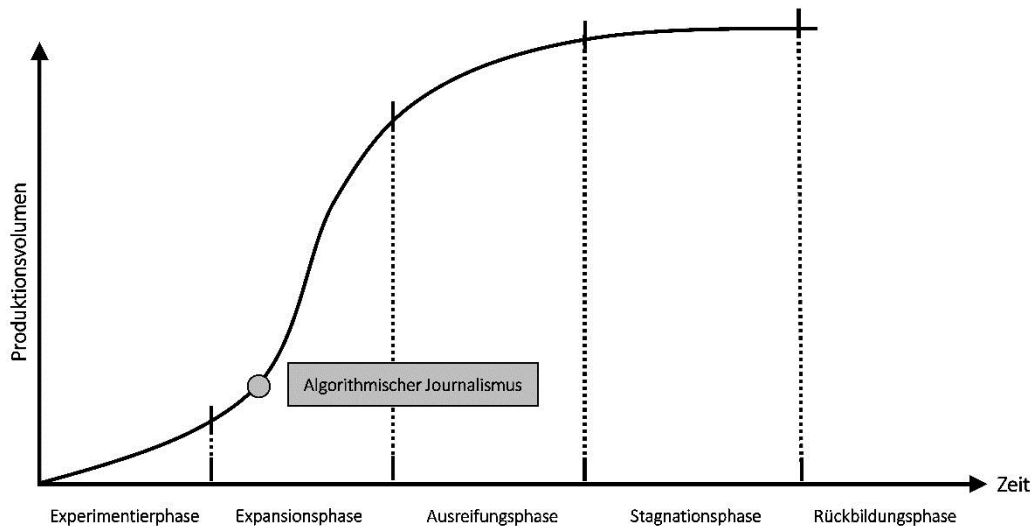
Das technische Potential von algorithmischem Journalismus kann dazu führen, dass mehr Inhalte zu spezifischen Themen schneller publiziert werden können (Artikel II und III). Dabei muss für jedes gewünschte Produkt eine eigene Syntax im Software-System erstellt bzw. der Algorithmus angepasst werden. Das Potential hat so Auswirkungen auf Arbeitsprozesse der Nachrichtenproduktion: Einerseits haben Journalisten mehr Zeit für andere Aufgaben, andererseits kann aber die Automatisierung auch zu Rationalisierungsmaßnahmen führen (vgl. van Dalen 2012; Artikel III). Dies hätte im Rahmen des SVE-Paradigmas Auswirkungen auf den Beschäftigungsgrad und das Marktergebnis. Besonders Nachrichtenagenturen wie Thomson Reuters oder Associated Press können in der Sport- oder Finanzberichterstattung neue Themenfelder erschließen und gegebenenfalls repetitive Aufgaben ersetzen (vgl. Artikel III; Anhang A4). Das hat wiederum Einfluss auf ökonomische Imperative der Medieninstitutionen.

Der zentrale Beitrag von Artikel II ist neben der technischen Analyse auch eine erste Übersicht über den Markt der automatisierten Textproduktion aus Anbieterperspektive und die Strukturierung seines Potentials in Marktphasen nach Heuss (1965). Mit den wenigen Anbietern automatisierter Inhaltsproduktion entsteht ein sachlicher Markt, der aufgrund der Möglichkeiten durch die Vielsprachigkeit der Software-Anwendungen aber nicht zwingend räumlich begrenzt ist (vgl. Artikel II). Die Anbieter fungieren hier als Layer Player. Sie verfügen über ein spezifisches Set von Ressourcen und Fähigkeiten bei der automatisierten Texterstellung. Aufgrund der Komplexität der Technik sind Medienorganisationen im Journalismus auf dieses Wissen angewiesen. Die Software-Anbieter können damit entscheidend zum Leistungskonzept (Value Proposition) des Journalismus beitragen, indem sie einerseits spezifische journalistische Produkte anbieten und andererseits mit dem Angebot der «Software as a Service» den Zugang zu der Technik ermöglichen. Sie liefern so den Möglichkeitsrahmen für die Leistungserstellung im Journalismus (vgl. Bieger & Reinhold 2011: 34ff.). Damit haben die Anbieter von algorithmischem Journalismus Einfluss auf die Ebene der Medienaussagen (Produkte) wie auch auf die Ebene der Mediakteure (Umgang mit der Technik).

Wie die Anbieterbefragung in Artikel II belegt, befindet sich der Markt in der Expansionsphase, da der Grad der Produktdifferenzierung und auch die Anzahl der Kunden

wachsen. Abbildung 2 zeigt das Marktphasenkonzept mit dem typischen Marktwachstumsverlauf von der Experimentier- bis zur Rückbildungsphase. In der Übersicht ist algorithmischer Journalismus entsprechend seinem derzeitigen Wachstumsverlauf der entsprechenden Marktphase zugeordnet.

Abbildung 3: Einordnung von AJ in Marktphasen in Anlehnung an Heuss (1965) und Latzer et al. (2014)



Eine Analyse der Marktstruktur auf Kundenseite – der Ebene von Medienorganisationen – ist wünschenswert, beinhaltet aber die Schwierigkeit, dass diese noch sehr zurückhaltend mit dem Thema «automatisierter Journalismus» umgehen und journalistische Inhalte – obschon bereits öffentlich – oft auch nicht als solche gekennzeichnet sind (vgl. Artikel II). Der Wille zur Transparenz ist je nach Organisation unterschiedlich ausgeprägt und beeinflusst damit insbesondere den ethischen Diskurs über die normative Qualität auf der Ebene der Mediensysteme.

Unklar bleibt weiterhin, ob und wie die Technik zur Beseitigung der finanziellen Schieflage im Journalismus beitragen kann und sich gegebenenfalls neue journalistische Geschäftsmodelle entwickeln oder bestehende angepasst werden müssen. Ein Thema, das mit der fortschreitenden Diffusion der Technik relevanter wird und im Rahmen der Dissertation auch aufgrund der Neuartigkeit des Phänomens nicht untersucht wurde. Dies sind optionale Fragestellungen für die zukünftige Forschung. In diesem Zusammenhang werden auch Fragen der Preispolitik bzw. Kostenstruktur relevant.

3. Einfluss auf die Ebene der Medienaussagen

In Bezug auf den Funktionskontext soll zunächst übergeordnet der Einfluss der Technik auf die journalistische Wirklichkeitskonstruktion diskutiert werden. Dabei geht es um die spezifischen Leistungen, die algorithmischer Journalismus übernehmen kann. Sowohl die

gesellschaftliche als auch die organisatorische und professionelle Ebene wirken auf die Ebene der Medienaussagen ein. Die Potentialanalyse von Artikel II stellt ebenjene Leistungen der Technik zur Ermöglichung gesellschaftlicher Kommunikation in den Vordergrund. Auch Artikel III thematisiert diese Leistungen der Technik in der Aussagenproduktion zusätzlich aus der Perspektive professioneller Akteure. Es zeigt sich, dass die Technik institutionelle Aufgaben des Journalismus erfüllen kann: Mit Verweis auf die eingeschränkte thematische Vielfalt (Universalität) und die Limitierungen bei der Kontextualisierung, Reflexion und Interpretation der digitalen Daten übernimmt die automatisierte Texterstellung Leistungen des Informationsjournalismus. Eingebunden in redaktionelle Strukturen professioneller Akteure kann die Technik nach erstmaliger Programmierung aktuelle und relevante Informationen zu bestimmten thematischen Schwerpunkten – z.B. Sport oder Finanzen – generieren und, falls programmiert, auch veröffentlichen. Wie Artikel I, II und III zeigen, lassen sich aktuell aber nur bestimmte und faktenbasierte Berichterstattungsmuster und Darstellungsformen verwirklichen. Dennoch erfüllt die Technik die Kriterien der Aktualität, Publizität, Periodizität und – mit Einschränkungen – auch der Universalität (vgl. Artikel II).

Aus der Perspektive der Journalisten werden das technische Potential und die Auswirkungen auf Arbeitsprozesse weitaus kritischer eingeschätzt. Die Ergebnisse von Artikel III beleuchten und erweitern die formal-technische Funktionsanalyse dabei zusätzlich aus der Anwenderperspektive professioneller Akteure. Denn es sind die Journalisten, die maßgeblich an der Wirklichkeitskonstruktion beteiligt sind. Wie Artikel III ergänzt, mangelt es algorithmischem Journalismus aus professionellem Blickwinkel u.a. an Kreativität, dem «human touch» oder dem Zwang, Schlagzeilen und Ereignisse im Vorfeld programmieren zu müssen (vgl. Artikel III). Dennoch wird auch sein Potential für die Berichterstattung erkannt. So könnten sich Möglichkeiten für eine objektivere Berichterstattung ergeben. Dies setzt aber sowohl vollständige Datensätze als auch eine objektive Programmierung voraus. Denn ein mögliches Bias kann bei Missachtung dieser Voraussetzungen auch im algorithmischen Journalismus entstehen. Die ethischen Dimensionen normativer Qualität wirken damit auch auf die Ebene des Funktionskontexts ein.

Vorteile der Technik lassen sich auch auf sprachlicher Ebene feststellen. So kann die Technik bei Anbietern wie AX Semantics Inhalte in über 12 verschiedenen Sprachen generieren (im Jahr 2015, vgl. Artikel II). Damit erweitert sich der Kreis eines möglichen Publikums, was wiederum zu Auswirkungen auf die Ebene der Medieninstitutionen und möglichen ökonomische Implikationen führt. Immer mehr rückt dabei der Leser mit seinen spezifischen Wünschen und Bedürfnissen in das Zentrum der Nachrichtenproduktion (vgl. Blöbaum 2000: 135). Für die zukünftige Entwicklung des algorithmischen Journalismus stellt sich dabei insbesondere auch die Frage, ob diese Wünsche erfüllt werden kön-

nen und ob und wie Texte gegebenenfalls personalisiert werden können. Wie erste Rezeptionsstudien zeigen, können Leser kaum einen Unterschied zwischen manuell und automatisiert erstellten Texten erkennen (vgl. Clerwall 2014; Graefe et al. 2016). Doch die Frage, ob sie diese Inhalte auch lesen, geschweige auch dafür bezahlen würden, ist noch nicht beantwortet. Denn damit steht und fällt auch unweigerlich der ökonomische Erfolg der automatisierten Textproduktion (Einfluss auf die Ebene der Medieninstitutionen).

Die Verständlichkeit und Qualität der automatisierten Texte sind von zwei Faktoren abhängig. Einerseits ist die Qualität der Software entscheidend. Mit «Qualität» ist hier das technische Potential der Software-Anwendungen bzw. Plattformen gemeint. Denn das technische Potential unterscheidet sich je nach Anbieter hinsichtlich Usability sowie spezifischer Optionen im Rahmen der NLG. Einige davon sind beispielsweise die Integration von Machine-Learning-Prozessen, die Nutzung externer Datensätze (Weltwissen) oder die sprachliche Vielfalt. Diese Unterschiede führen auch dazu, dass Anbieter unterschiedliche Produkte anbieten können bzw. journalistische Kunden das Potential der Technik selbstständig und vielfältiger nutzen können (vgl. Artikel II und III). Andererseits hängt die Qualität der Medienaussagen auch von den Journalisten, deren Rollenkonzeptionen und deren technischen Fähigkeiten bei der Nutzung von algorithmischem Journalismus ab (vgl. Artikel III). Beide Akteure haben damit direkten Einfluss auf die Aussagenproduktion. Folglich stehen die organisatorische und die professionelle Ebene mit der Ebene der Medienaussagen in Wechselwirkung. Dabei findet die journalistische Wirklichkeitskonstruktion von algorithmischem Journalismus nur in sozialen Prozessen und im Wechselspiel mit diesen Akteuren statt. Nur innerhalb dieses Rahmens kann die Technik ihre regelnde und steuernde Kraft entfalten. Verwiesen sei hier zusätzlich auf die Rolle des Internets, das aufgrund seiner Architektur die Nutzung und Verbreitung von algorithmischem Journalismus überhaupt erst ermöglicht. Neben den Möglichkeiten beim Zugang und bei der Distribution der generierten Texte ist das Internet der Datenlieferant für die automatisierte Textproduktion. Artikel I, II und III thematisieren auch die Rolle von Big Data im algorithmischen Journalismus. Daten sind der Rohstoff für die automatisierte Textproduktion. Medienorganisationen sind dabei für die Datenerhebung selbst verantwortlich (siehe Abschnitt 1: *Einfluss auf die Ebene der Mediensysteme*). Forschungsbedarf besteht hier insbesondere in Bezug auf die Rolle externer Datenanbieter (Data Broker) und auch in Bezug auf Prozesse der Datensammlung, -speicherung und -aufbereitung in Medienorganisationen. Denn je vielfältiger der Datensatz ist, desto mehr inhaltliche Optionen ergeben sich auch für die Aussagenproduktion. Die Rolle der Daten als Quelle der automatisierten Texterstellung ist noch nicht hinreichend untersucht. Welche Quellen nutzen die Medienorganisationen als Grundlage? Wie erheben sie die Daten? Wie werden die Daten hinsichtlich normativer Kriterien geprüft? Beziehen sie die Daten von externen Daten Providern oder sind es Daten, die öffentlich verfügbar sind? Speziell bei Fragen des Zugangs ergeben sich

Auswirkungen auf die Vielfalt der automatisierten Texte. Denn dann, wenn die Daten öffentlich verfügbar sind (z.B. Open Government Data), hat prinzipiell jeder journalistische Anbieter die Möglichkeit, diese Daten auch zu verarbeiten. Dies führt dazu, dass inhaltlich ähnliche Texte in großer Stückzahl produziert werden, die sich nur sprachlich unterscheiden. Dies wiederum führt zu ökonomischen Fragen auf organisatorischer Ebene. Wie können Medienorganisationen einen inhaltlichen und ökonomischen Mehrwert schaffen? Welches ist das Alleinstellungsmerkmal für die automatisierten Texte? Wie kann man sich mit selbstständig erhobenen oder extern zugekauften Daten von der Konkurrenz absetzen? Diese Fragen und Entscheidungen bezüglich des Datenzugangs und der Datenauswahl haben in letzter Instanz Auswirkungen auf die inhaltliche Vielfalt wie auch auf Geschäftsmodelle im Medienmarkt.

4. Einfluss auf die Ebene der Medienakteure

Abschließend soll der Einfluss von algorithmischem Journalismus auf die Ebene der Medienakteure mit Rückgriff auf das Konzept der gradualisierten Handlungsträgerschaft von Rammert & Schulz-Schaeffer (2002) diskutiert werden (siehe Artikel I, II und III). Neben der Rolle der Journalisten und des Algorithmus wird auch die Rolle der Software-Anbieter in die Diskussion einbezogen. Mit der Potentialanalyse in Artikel II kann das Wechselspiel zwischen Journalist und Algorithmus sichtbar gemacht werden. Dabei geht es «nicht um Prognosen oder Imperative, sondern um vielfältige Optionen, die verhindert oder ausgewählt und institutionalisiert werden können» (Neuberger 2001: 92).

Artikel I zeigt, dass algorithmischer Journalismus im Vergleich zu Computer-Assisted Reporting oder Data Journalism die höchsten Freiheitsgrade bei der Produktion von Nachrichten bietet. Es ergeben sich unterschiedliche Interaktivitätslevel zwischen Journalist und Algorithmus.

Im Sinne der Operationalisierung von Rammert und Schulz-Schaeffer (2002) übernimmt algorithmischer Journalismus institutionelle Aufgaben des Journalismus bei der Beobachtung und Beschreibung der Gesellschaft. Ihm kann eine Form von Handlungsträgerschaft zugewiesen werden (zur Frage der Verantwortung siehe Artikel V und Abschnitt 1: *Einfluss auf die Ebene der Mediensysteme*; Just & Latzer 2016: 252). Mit Verweis auf das E-V-A-Modell ist auf der Ebene der Eingabe (E) ein schwaches Interaktionslevel zu identifizieren (Level 1: *verändernde Wirksamkeit*). Der Algorithmus führt hier die Eingaben des Journalisten aus. Dabei agiert er strikt nach den Regeln der Eingabeaufforderung des menschlichen Akteurs. Die Auswahl der Datenbasis, die Festlegung des Themas, die Programmierung der Struktur bzw. Schablone für die zu automatisierenden Texte (Sprache und Verständlichkeit) sowie die Fixierung von Ort und Zeit der Publikation sind Entscheidungen, die der Journalist/die Redaktion im Vorfeld definieren kann und muss. Damit hat er die

Kontrolle über die Leistungen der Aktualität, Universalität, Periodizität und Publizität der automatisierten Texte sowie indirekt auch über den Zugang des Publikums zu diesen Inhalten (vgl. Artikel I und II). Dies trifft aber nur zu, solange die Inhalte auf der eigenen Plattform publiziert werden, z.B. auf einem Online-Nachrichtenportal. Werden diese Inhalte über andere reichweitenstarke Kanäle wie z.B. Facebook distribuiert, kann automatisierter Journalismus zwar ein heterogen zusammengesetztes Laienpublikum erreichen, jedoch haben hier Filter- und Empfehlungs-Algorithmen, wie z.B. der Facebook-News-Feed-Algorithmus, und indirekt auch das Publikum Einfluss auf die Verteilung der Nachrichten (vgl. Vos & Heinderyckx 2015). Die Abhängigkeit des Journalismus von externen Akteuren steigt damit an, sobald man die technische Architektur der Redaktion bzw. Organisation verlässt (vgl. Anhang A3). Problematisch ist diese Ebene, weil sich aktuell nur jene Textgattungen automatisieren lassen, die zuvor als Schemata strukturell festgelegt wurden und hauptsächlich auf Fakten basieren (vgl. Artikel I und II). Nennen lassen sich hier Beispiele aus der Wetter-, Sport- oder Finanzberichterstattung. Nur wenn ausreichend maschinenlesbare Fakten vorliegen, ist der Algorithmus in der Lage, diese Informationen auch zu extrahieren.

Innerhalb dieses festgelegten Regelwerks kann der Algorithmus auf der Ebene der Verarbeitung (V) agieren (vgl. Rammert & Schulz-Schaeffer 2002: 44f.). Hier zeigen sich in Bezug auf die Modulation der Sprache, aber auch bei der Kontextualisierung und Beschreibung von Inhalten, dass Algorithmen unabhängig von menschlicher Einflussnahme Auch-anders-handeln-Können und eine Eigendynamik entwickeln (Level 2). Diese Fähigkeit umfasst zwei Elemente: Der Algorithmus kann auf seine Umwelt eingehen und dabei seine Funktion auch bei Veränderung der Rahmenbedingungen weiter erfüllen. Er kann sein eigenes Verhalten verändern und so unberechenbar werden (vgl. Rammert & Schulz-Schaeffer 2002: 45f.). Während der Journalist keinen unmittelbaren Einfluss auf die Ebene der Verarbeitung hat, zeigen sich gewisse Freiheitsgrade bei der Reihenfolge und in der Art und Weise, wie die Eingabeaufforderung des Journalisten umgesetzt wird. Je nachdem wie vielfältig das zugrunde liegende sprachliche Regelwerk ausfällt, kann der Algorithmus auf dieser Ebene beispielsweise Synonyme für Wortgruppen erstellen (z.B. schreibt er statt «groß» «riesig» oder «gewaltig»), beim Satzanfang variieren oder andere Adjektive verwenden. Mit neuen Ansätzen des Machine Learnings erkennt der Algorithmus Muster und Gesetzmäßigkeiten in seinen Texten und kann so auch eine gewisse Vielfalt erzeugen, eine solide Datenbasis vorausgesetzt. Im Rahmen der menschlich erzeugten Regelwerke kann die Technik hier selbstständig entscheiden.

Die Möglichkeit des Auch-anders-handeln-Könnens zeigt sich insbesondere beim Rückgriff auf das Konzept des Weltwissens in der Computerlinguistik. Weltwissen – im Vergleich zu sprachlichem Wissen – ist das Wissen über Erfahrungen und Kenntnisse aus der Umwelt und Gesellschaft: Man weiß z.B., dass man nicht mit Feuer spielen sollte, weil

man sich dabei verbrennen kann. Wenn dieses Weltwissen – sei es auch nur für bestimmte Bereiche wie z.B. Fußball – codiert wird, ergeben sich neue Handlungsspielräume (Freiheitsgrade) für Algorithmen bei der Kontextualisierung von Informationen. Aktuell befindet sich diese Codierung in einem Frühstadium (vgl. Artikel III). Dennoch ist es das erklärte Ziel der Computerlinguistik, Weltwissen – faktisches Weltwissen und andere Wissensformen wie episodisches Wissen und Erfahrungen – Schritt für Schritt digital codieren zu können (vgl. Stede, Anhang A1). Sollte dies in Zukunft für weitere Themenfelder möglich werden, würde der Algorithmus hier die zentrale Leistung des Journalisten bei der Kontextualisierung von Informationen übernehmen, ist es doch gerade die Einordnung von Wissen in gesellschaftliche Zusammenhänge, die als zentrale Leistung bis jetzt nur «menschlichen» Journalisten zugeschrieben werden kann.

Auf der Ebene der Ausgabe (A) zeigt sich einerseits ein schwaches Interaktionslevel (Level 1) in Bezug auf Publikation und Distribution der Texte, auch weil zentrale Merkmale des Texts bereits bei der Eingabe festgelegt werden. Andererseits lässt sich hier auch ein mittleres Interaktivitätslevel mit Rückgriff auf Machine-Learning-Prozesse deuten. Algorithmen können ihre Texte selbstständig verbessern, wenn sie mit einer Vielzahl von Texten «gefüttert» werden. Dabei können sie Muster erkennen und diese bewerten, um einen Idealtext auf Grundlage der Daten zu produzieren und diesen dann zu variieren.

Diese Differenzierung der technischen Funktionsweise in Artikel I und II zeigt auch, dass der menschliche Akteur, eingebettet in organisationale Strukturen der Redaktion, noch immer die Hoheit über die Aussagenproduktion hat und damit auf die Ebene der Medienaussagen einwirkt. Seine regelnde und steuernde Kraft gewinnt algorithmischer Journalismus daher nur in Wechselbeziehung mit menschlichen Akteuren und seiner Distribution auf reichweitenstarken Kanälen. Er kann hier als Werkzeug operationalisiert werden, das zwar selbstständig innerhalb bestimmter Regelwerke handeln kann, sein Handlungsspielraum ist – wie auch die Journalisten in Artikel III feststellen – aber stark durch seine technische Disposition eingeschränkt und von externen Faktoren wie der Qualität der Software-Plattform, der Programmierung oder der Qualität der Daten abhängig (vgl. Ausführungen zu Abschnitt 3: *Einfluss auf die Ebene der Medienaussagen*). Artikel III zeigt auch, dass es zu Funktionsproblemen bei der Technik kommen kann, wenn Normen und Standards aufgrund fehlender Kenntnisse nicht eingehalten werden. Die Befunde in Artikel III können damit einen Hinweis geben, welche Fähigkeiten im algorithmischen Journalismus relevant werden können. Essentiell ist dabei die Fähigkeit, strukturell-analytisch und mathematisch Denken zu können. Diese wird sowohl bei der Programmierung der Syntax wie auch bei der Analyse der Datensätze immer wichtiger. Konkrete Programmierkenntnisse, wie z.B. Python, sind keine elementare Voraussetzung, wenn die Technik auf redaktioneller Ebene genutzt werden soll. Denn die Software-Anbieter verbessern ihre Plattfor-

men kontinuierlich im Hinblick auf die Nutzbarkeit. Medienakteure wirken auf die organisatorische Ebene ein, indem sie im Austausch mit den Software-Anbietern die Entwicklung auf diesem Gebiet vorantreiben.

Eng verknüpft mit der Rolle der Journalisten und der Technik sind auch die Rolle der Software-Anbieter und Programmierer im automatisierten Journalismus sowie die Frage, ob sie professionelle Akteure im Journalismus sind. Denn mit dieser Entscheidung stellen sich insbesondere auch Fragen der Verantwortung und normativen Qualität (vgl. Abschnitt 1: *Einfluss auf die Ebene der Mediensysteme*). Mithilfe verschiedener formaler und inhaltlicher Kriterien können sich journalistische Organisationen und Akteure von nicht-journalistischen unterscheiden lassen. Folgt man der Definition von Weischenberg et al. (2006b: 347) ist man Journalist, wenn man hauptberuflich in fest angestellter oder freier Mitarbeit mit der Produktion journalistischer Medienangebote verknüpft ist. Hauptberuflichkeit wird dann attestiert, wenn ein Journalist mehr als die Hälfte seiner Einkünfte aus journalistischer Arbeit bezieht oder mit mehr als der Hälfte seiner Arbeitszeit für journalistische Medien tätig ist. Ausgeschlossen haben die Autoren Arbeitsrollen, die keinen unmittelbaren Einfluss auf den Inhalt der redaktionellen Produkte haben. Ehrenamtliche, arbeitslose oder nebenberuflich freie Journalisten werden in ihrer Studie somit nicht als professionelle Akteure des Journalismus berücksichtigt. Eine abschließende Bewertung des Akteursstatus der Software-Anbieter kann aufgrund mangelnder empirischer Befunde nicht abgegeben werden. Dennoch lassen sich beispielsweise einige Faktoren nennen, die für eine zukünftige Diskussion relevant sein können:

Variante 1 (kein journalistischer Akteur): Die redaktionelle Entscheidung (Datenauswahl), welche Inhalte automatisiert werden, wird nicht von den Anbietern getroffen, sondern liegt in der Entscheidungsgewalt der Medienmacher und Medienbesitzer. Diese sind es, die die Textschablone programmieren und so die inhaltliche Struktur der Texte bestimmen. Anbieter liefern das technische Werkzeug.

Variante 2 (journalistischer Akteur/Zulieferer): Wenn allerdings Software-Anbieter auf Kundenwunsch die Texte programmieren und mit den Daten der journalistischen Organisationen arbeiten, ist die Abgrenzung schwieriger. Denn mit der Programmierung des Codes tragen sie in Abstimmung mit den Medienmachern und Medienbesitzern unmittelbar zur inhaltlichen Ausgestaltung der Medienprodukte bei. Dann stellt sich die Frage, ob die Software-Anbieter mit mehr als der Hälfte ihrer Einkünfte aus dieser Arbeit beziehen oder mit mehr als der Hälfte ihrer Arbeitszeit für journalistische Medien tätig sind. Eine Frage, die noch nicht eindeutig geklärt ist. Denn Software-Anbieter haben neben journalistischen Kunden oft auch weitere, meist finanziell potentere Kunden in anderen Industriezweigen, z.B. im E-Commerce. Eine Einschätzung, die nicht nur Tom Kent von Associated Press teilt: «[...] I don't think it's [Journalism: KD] necessarily the core of their work. In fact, we do a lot of work for them and help them to train these arguments» (Kent, Anhang

A4). Mit ihrem Partner Automated Insights ist die Nachrichtenagentur Vorreiter auf dem Gebiet des algorithmischen Journalismus. Zukünftig müsste für eine Einschätzung aber je nach Anbieter und auch im Hinblick auf dessen individuelles Produkt- und Leistungsangebot unterschieden werden.

Eine abschließende Bewertung sollte auf weitere empirische Kenntnisse zurückgreifen und Kriterien wie die Selbsteinschätzung der Software-Anbieter und die Einschätzung von Medienmachern und Medienbesitzern sowie Fragen zu Kriterien der Sicherung normativer Qualität einbeziehen. Erste Indizien lassen erkennen, dass die Anbieter sich dieser Verantwortung entziehen wollen und nur noch den Zugang zu Software und Schulungen in diesem Rahmen anbieten (vgl. Artikel III).

5. Zusammenfassung

Mit Rückgriff auf das Konzept von Technik als Akteur und Institution sowie der Einbettung der Anwendung in das Forschungsprogramm «Algorithmische Selektion im Internet» konnte die Rolle des algorithmischen Journalismus als zentrale Triebkraft im Medienwandel identifiziert und analysiert werden. Um seine Rolle für kommunikationswissenschaftliche Fragestellungen anschlussfähig zu machen, wurde das Kontext-Modell von Weischenberg (1992) herangezogen. Dieses Modell ermöglicht es, den Einfluss der Technik auf journalistisches Handeln theoretisch wie empirisch greifbar zu machen. Die Dissertation liefert mit der Kombination soziologischer und kommunikationswissenschaftlicher Perspektiven somit einen zentralen Mehrwert für die Analyse des Medienwandels und für zukünftige Forschungsvorhaben bei der Integration der Rolle von Technik. Mit Verweis auf die übergeordnete Forschungsfrage zeigen erste theoretische und empirische Befunde der Artikel I bis V, welchen Einfluss algorithmischer Journalismus auf gesellschaftlicher, organisatorischer und professioneller Ebene des Journalismus hat und welche Interdependenzen sich als Konsequenz zwischen diesen Ebenen ergeben. Die Dissertation betreibt Grundlagenforschung, wohl wissend, dass sie damit nur einen ersten Beitrag dazu leisten kann, das Forschungsfeld weiter zu systematisieren und einer kommunikationswissenschaftlichen Analyse zugänglich zu machen.

Die Arbeit verortet die ethischen Herausforderungen des algorithmischen Journalismus dabei auf professioneller Ebene sowie der Ebene einer individuellen Ethik des Journalisten und bezieht auch die strukturellen Rahmenbedingungen des journalistischen Handlungsfeldes auf organisatorischer und gesellschaftlicher Ebene in die Analyse ein. Auf der Ebene der Mediensysteme zeigt sich, dass algorithmischer Journalismus die normativen Anforderungen der Nachrichtenproduktion für die Akteure der professionellen und organisatorischen Ebene beeinflusst und die Verantwortungsebenen vor, während und nach der Nachrichtenproduktion neu regelt. Dies bezieht sich sowohl auf Fragen der

Auswahl und Nutzung von Daten als auch auf die Prozesse der Verarbeitung und die Publikation der Texte. Unklar bleibt, wie Medienmacher und Medienbesitzer generell mit diesen Herausforderungen von Technik und Big Data umgehen und ob diese in letzter Instanz auch zu angepassten Ethik-Kodizes führen, z.B. in den übergeordneten «Publizistischen Grundsätzen» oder auch auf organisatorischer und redaktioneller Ebene in Form von ethischen Richtlinien und Standards für die Nachrichtenproduktion. Denn die technische Entwicklung führt dazu, dass Handlungsorientierungen und -anweisungen für den Umgang mit und die Produktion von journalistischen Inhalten auf organisatorischer wie professioneller Ebene kontinuierlich angepasst werden müssen. Das Konzept der gradualisierten Handlungsträgerschaft und die Technikpotentialanalyse ermöglichen es, die Ebenen der Verantwortungen zu identifizieren und die Rolle der Technik einer breiteren normativen Analyse zugänglich zu machen. Ethische Fragestellungen zur Transparenz wurden aus der Perspektive professioneller Akteure in Artikel III sichtbar und sollten spezifisch qualitativ und quantitativ auf der Ebene der Medieninstitutionen wie auch der Medienakteure untersucht werden. Wiesenhof, Zerfass und Moreno (2017) tun dies beispielsweise für den Bereich der strategischen Kommunikation und der Public Relations. Insbesondere Artikel III zeigt, dass innerhalb des normativen Rahmens ethische Herausforderungen des algorithmischen Journalismus sowohl individuell als auch auf organisatorischer Ebene unterschiedlich interpretiert werden (vgl. auch Anhang A4; zur Kennzeichnung der Inhalte vgl. Reich 2010; Montal & Reich 2016). Hier besteht weiterer Forschungsbedarf.

Auf der Ebene der Medieninstitutionen ergeben sich durch den Einfluss der automatisierten Textproduktion organisatorische, technische und ökonomische Implikationen für das journalistische Handeln. Diese Faktoren haben wiederum Auswirkungen auf die Ebene der Medienaussagen und die Ebene der Medienakteure. Es verändern sich nicht nur Arbeitsprozesse und Strukturen in der Nachrichtenproduktion auf redaktioneller Ebene, sondern die Technik hat auch das Potential, Marktstrukturen und Geschäftsmodelle des Journalismus zu verändern. Insbesondere hier besteht weiterer Forschungsbedarf, denn die Ergebnisse der Artikel II und III liefern mit der Einordnung des algorithmischen Journalismus in Marktphasen nur erste Hinweise hinsichtlich einer Diffusion der Technik im Nachrichtenmarkt. Das technische Potential kann zudem zu Spar- und Rationalisierungsmaßnahmen führen, die insbesondere die Ebene der Medienakteure treffen würden (Artikel I, II und III).

Auf der Ebene der Medienaussagen hat die automatisierte Textproduktion unmittelbaren Einfluss darauf, wie Inhalte gestaltet werden können, und damit auch darauf, welche Aufgaben des Informationsjournalismus algorithmischer Journalismus bei der Beobachtung und Beschreibung der Gesellschaft übernimmt. Sowohl die Ebene der Medien-

institutionen als auch die Ebene der Medienakteure stehen mit der Ebene der Medienausagen in enger Beziehung und erbringen diese Leistungen gemeinschaftlich mit der Technik.

Auf der Ebene der Medienakteure ergeben sich unterschiedliche Interaktivitätslevel zwischen Journalist und Algorithmus in der Nachrichtenproduktion. Der Algorithmus hat Freiheitsgrade, also einen Gestaltungsspielraum, bei der Verarbeitung der Daten, wenngleich die finale Kontrolle über Daten, Themen und Struktur der Texte in «menschlicher» Hand verbleibt. Mit dem Rückgriff auf techniksoziologische Konzepte könnte insbesondere die Journalismusforschung profitieren und ihren bislang auf Menschen oder Organisationen bezogenen Akteursbegriff erweitern.

Aus ethischer Perspektive sind es die Medienmacher und Medienbesitzer, die sich an professionsethischen Standards der organisatorischen und gesellschaftlichen Ebene orientieren und als Träger der Verantwortung im algorithmischen Journalismus zu identifizieren sind (Artikel IV und V). Die spezifische Ausgestaltung und Berücksichtigung normativer Qualitätskriterien in der Nachrichtenproduktion ist auf organisatorischer und professioneller Ebene nicht einheitlich ausgeprägt und befindet sich in einer Adaptionsphase. Die abschließende Bewertung der Rolle der Software-Anbieter als journalistische Akteure muss jedoch erst durch weitere empirische Befunde erhärtet werden und ist auch abhängig von der zukünftigen strategischen Ausrichtung der Anbieter.

Kapitel V diskutiert die skizzierten offenen Fragestellungen noch einmal übergeordnet in Bezug auf die Ebenen des Kontext-Modells und gibt Implikationen für die zukünftige Forschung.

V. Abschließende Betrachtung und Forschungsausblick

Der wissenschaftliche Beitrag der Dissertation liegt in der explorativen Analyse und Strukturierung des algorithmischen Journalismus im Forschungsfeld des Computational Journalism (vgl. Artikel I). Dabei werden die Einflüsse der Technik auf journalistisches Handeln auf gesellschaftlicher, organisatorischer und professioneller Ebene identifiziert und diskutiert (vgl. Tabelle 1). Wie in Kapitel IV gezeigt wurde, verlaufen diese in einem stetigen Wechselspiel, ohne dass eine eindeutige Verlaufsrichtung festgestellt werden kann. Denn algorithmischer Journalismus beeinflusst nicht nur jede Ebene des Kontext-Modells, sondern führt auch dazu, dass sich Anpassungsprozesse zwischen diesen abzeichnen. Dabei entscheidet sich die gesellschaftliche Wirkung von algorithmischem Journalismus letztlich durch das koevolutionäre Zusammenspiel mit diesen Ebenen. Die Innovation «algorithmischer Journalismus» dient dabei als Beobachtungsreferenz, um die Dynamiken des medialen Wandels im Journalismus sichtbar zu machen (vgl. Dogruel 2013). Doch wie dieser Verlauf konkret voranschreiten wird und welche Konsequenzen sich für den Journalismus als professionelles Beobachtersystem ergeben, ist aufgrund der rasanten technischen Entwicklung und der immer vielfältigeren Möglichkeiten beim Zugang zu und bei der Speicherung von Big Data ungewiss (vgl. Weischenberg 1995: 537). Der explorative Charakter der Studie ist dabei der Neuartigkeit des Phänomens und dessen Einsatz im Journalismus geschuldet. Die Dissertation kann damit nur erste Ergebnisse liefern, um die komplementäre und integrative Beziehung der Technik im Journalismus einer breiteren Analyse zugänglich zu machen. An dieser Stelle sollen mit Bezug auf die Ebenen des Kontext-Modells weitere offene Fragestellungen thematisiert werden.

Eingebettet ist die zukünftige Entwicklung von algorithmischem Journalismus in den sozio-technischen Kontext des Internets als Informations- und Kommunikationstechnologie (vgl. Sandvig 2013). Denn seine Architektur erleichtert für Medienmacher und Medienbesitzer den Zugang zu Big Data sowie zur Technik selbst und ermöglicht die Distribution der Texte über öffentliche Kanäle. Mit algorithmischem Journalismus wächst dabei die verfügbare Informationsmenge, auch weil die Informationen im Internet räumlich und zeitlich fast unbegrenzt abrufbar sind. Zwar wurde gezeigt, dass algorithmischer Journalismus zur Komplexitätsreduktion beitragen kann, indem er digitale Daten aufbereitet und damit weitere Themenfelder erschließt (vgl. Artikel I, II und III; Erdmann, Anhang A3). Doch mit dieser Entwicklung stellt sich auch die Frage, ob die Technologie Komplexität womöglich nicht nur reduziert, sondern auch addiert, indem die Menge an Nachrichten, die algorithmischer Journalismus erstellt und über das Internet distribuiert, stetig zunimmt. Dieses Überangebot macht es nicht nur Journalisten schwerer, ihrer Gatekeeper-Rolle bei der Selektion der Daten und möglicher Themenfelder gerecht zu werden, sondern auch das Publikum muss aus diesem zusätzlichen Angebot selektieren.

Die Frage, inwieweit professionelles journalistisches Handeln durch algorithmischen Journalismus weiter automatisiert und substituiert werden kann, hängt dabei von zwei Faktoren ab: erstens davon, wie sich die Leistungsfähigkeit des algorithmischen Journalismus entwickelt. Dies entscheidet sich mit dem Zugang zu, der Sammlung, Aufbereitung und der Nutzung von Big Data durch Medienorganisationen sowie durch den technischen Fortschritt in der natürlichsprachlichen Textgenerierung¹⁶ (vgl. AX Semantics 2017; Artikel III). Von diesen Variablen hängen die möglichen inhaltlichen und themenspezifischen Gestaltungsräume der Technik im Journalismus ab. Dabei muss die Funktionserwartung des Journalismus – nämlich die Gesellschaft zu beschreiben und zu beobachten – auch auf algorithmischen Journalismus übertragen werden. Wie Artikel I, II und III zeigen, können menschliche Leistungen bereits für bestimmte Themenbereiche algorithmisch ergänzt, automatisiert und substituiert werden. Insbesondere die Potentialanalyse in Artikel II ermöglicht es, den Untersuchungsgegenstand auch für mögliche rechtliche (z.B. Haftung) und politische Fragestellungen (z.B. Regulierung externer Anbieter) zu öffnen.

Zweitens hängt die Frage einer weiteren Automatisierung und möglichen Substituierung menschlicher Arbeit auch vom Publikum und seinem Vertrauen in diese Inhalte ab (vgl. Henry 2007; Kuhlen 1999: 85ff.; Lewis 2011). Denn während u.a. die Leistungsfähigkeit der automatisierten Textproduktion Gegenstand dieser Arbeit war, ist die Rolle des Publikums sowohl in dieser Arbeit als auch in der übrigen Forschung noch unterrepräsentiert. Fragen der Nutzung der automatisierten Inhalte wurden bis jetzt ausschließlich im Hinblick auf die Qualität der Texte und ihre Unterscheidbarkeit von menschlich erzeugten Texten diskutiert (vgl. Clerwall 2014; Graefe et al. 2016). Das Publikum ist für zukünftige Forschungsvorhaben daher unbedingt einzubeziehen. Denn «[i]n der intelligenten Einbindung des Publikums könnte die Zukunft des professionellen Journalismus liegen» (Neuberger 2012: 15). Dabei geht es sowohl darum, zu wissen, wer das spezifische Publikum für algorithmischen Journalismus ist, als auch dessen Wünsche zu erkennen und auf diese zu reagieren (vgl. Hohlfeld 2005, 2016; Loosen 2016). Fragen zu inhaltlichen Spielräumen bei der Personalisierung automatisierter Nachrichten sind auf professioneller und organisatorischer Ebene ebenso zu nennen wie weitere Möglichkeiten der Nutzer-Beteiligung, z.B. bei der Datensammlung. Problematisch wird das nur, wenn sich die journalistische Themensetzung und Selektionsleistung ausschließlich am Leser orientieren und sich damit Leistungs- und Funktionsmerkmale des Journalismus verschieben. Die Rolle des Journalisten wird insbesondere bei der Datenanalyse und Kontextualisierung der Inhalte auf professioneller Ebene immer wichtiger. Denn das tatsächliche «Schreiben» eines Artikels verschiebt sich im algorithmischen Journalismus aktuell mit der Programmierung der Codezeilen auf die Ebene der Eingabe. Entscheidend werden auf dieser Ebene die Analyse von

¹⁶ Dies umfasst sowohl die Codierung von faktischem Wissen als auch die Codierung von anderen Wissensformen, z.B. episodischem Wissen oder Erfahrungswissen (vgl. Stede, Anhang A1).

Datensätzen mit Hilfe weiterer Software-Anwendungen sowie Fragen des Storytellings, also mit welchen Mitteln die journalistische Geschichte dem Publikum erzählt wird. Durch die Automatisierung des Schreibprozesses und der Publikationen bzw. seine Verlagerung könnten zudem der direkte Kontakt und die inhaltliche Auseinandersetzung des Journalisten mit dem Publikum verloren gehen. Die Wünsche des Publikums können zudem nicht während des Schreibprozesses adressiert werden, sondern müssen im Vorfeld analysiert und für das jeweilige journalistische Produkt entsprechend programmiert werden (vgl. Sola Pool & Shulman 1959).

Wie algorithmischer Journalismus zum Aufbau oder Erhalt von Vertrauen in den Journalismus beitragen kann, ist ebenfalls noch unbeantwortet (vgl. zum Datenjournalismus Possler & Link 2017). Denn während Vertrauen bisher maßgeblich durch die Themenselektion sowie die Beschreibung und Bewertung der Inhalte durch den Journalisten geprägt wurde (vgl. Matthes & Kohring 2003), muss das Publikum im algorithmischen Journalismus Vertrauen in die Selektionsleistungen des Algorithmus entwickeln. Das Vertrauen in Daten sowie das Vertrauen in die korrekte Programmierung der Syntax durch die Journalisten und die Plattform der externen Anbieter werden hierbei vorausgesetzt. Umso offensiver müsste mit dem Gebot der Transparenz bei der Kennzeichnung von algorithmischem Journalismus auf organisatorischer und professioneller Ebene umgegangen werden (vgl. Anderson 2011a; Diakopoulos 2015; Artikel IV und V). Diese Vertrauensbildung ist mehrdimensional und abhängig sowohl von algorithmischen Leistungen anderer Anwendungen und Plattformen, z.B. Recommender-Systeme auf Facebook, als auch von anderen Akteuren (Anbietern/Freunden/Publikum), die sich auf diesen Plattformen bewegen (vgl. Anderson 2012; Napoli 2014). Vom algorithmischen Journalismus muss daher auch erwartet werden, dass er diese Machtstrukturen sichtbar macht und normative Fragestellungen der Transparenz und Objektivität adressiert (vgl. Anderson 2011a; Neuberger & Nuernbergk 2015). Hier liegt auch sein Potential: Objektive Programmierung und vollständige Datensätze vorausgesetzt, bildet er automatisiert inhaltlich das ab, was im Programmcode intendiert war. Er könnte damit zu einer umfassenderen und präziseren Wirklichkeitsdarstellung und zur Bildung von Vertrauen beitragen.

Inwieweit die automatisierte Texterstellung in das redaktionelle Umfeld auf organisatorischer Ebene bereits integriert ist und sein wird und wie sich der Nachrichtenmarkt und Geschäftsmodelle des Journalismus durch den Einfluss der Technologie entwickeln werden hängt nicht zuletzt auch von Entscheidungen im Medienmanagement ab. Die Entscheidung, ob algorithmischer Journalismus genutzt wird, muss immer individuell auf die Medienorganisation, deren Marke und das Leistungsportfolio abgestimmt sein (vgl. Wagner, Anhang A2). Hier stellt sich insbesondere die Frage, ob algorithmischer Journalismus zur Befreiung der Journalisten von Routineaufgaben eingesetzt wird oder ob die Technik

in größerem Ausmaß zur Substituierung und Rationalisierung menschlicher Arbeit beiträgt (vgl. Artikel III). Erste Hinweise auf den Markt und mögliche Rationalisierungsmaßnahmen des algorithmischen Journalismus konnten mit der Anbieter- und Journalistenbefragung in Artikel II und III gegeben werden.

Ethische Fragestellungen zum Umgang mit algorithmischen Journalismus, zur Transparenz und Verantwortung müssen sowohl auf organisatorischer als auch auf professioneller Ebene umfassend adressiert werden (vgl. Diakopoulos & Koliska 2016; Montal & Reich 2016). In Artikel IV und V konnten die ethischen Herausforderungen der Technik verortet und Fragen zur normativen Qualität aus der Perspektive der journalistischen Ethik diskutiert werden. Dabei sind auch die Rolle von Big Data sowie die Rolle von Datenlieferanten und Software-Anbietern in die zukünftige Forschung zu algorithmischem Journalismus einzubeziehen.

Der Umgang mit der automatisierten Textproduktion und deren technisches Potential werfen zudem Fragen der journalistischen Ausbildung auf. Wie muss auf professioneller und organisatorischer Ebene in Medienorganisationen und auf Seiten der universitären und hochschulgebundenen Ausbildung damit umgegangen werden? Welche technischen, methodischen und analytischen Fähigkeiten sind relevant und müssen in das Curriculum aufgenommen werden (vgl. Dernbach 2016; Artikel III)?

Die Liste weiterer Fragestellungen ist lang, auch weil die technische Entwicklung, die organisationale Integration und Auseinandersetzung mit dem Phänomen sowie die Forschung zu algorithmischem Journalismus noch in den Startlöchern stecken. Daher wurden hier nur einige mögliche Forschungsschwerpunkte skizziert. Die zukünftige Forschung sollte algorithmischen Journalismus nicht als Einzelphänomen analysieren. Denn dessen gesellschaftliche Wirkkraft potenziert sich im Zusammenspiel mit den vielfältigen technischen Möglichkeiten und Anwendungen des Computational Journalism. Diese reichen von der Datensammlung über die Datenaufbereitung (Visualisierung) bis zur Distribution der Inhalte.

Am Ende geht es nicht um den Kampf von Mensch gegen Maschine, sondern «viel eher um deren komplementäre und integrative Beziehung» (Neuberger & Nuernbergk 2015: 216). Die Dissertation trägt einen kleinen Teil dazu bei, dieses Wechselspiel auf gesellschaftlicher, organisatorischer und professioneller Ebene im Journalismus sichtbar zu machen.

VI. Publikationen

Artikel I

Algorithmen, Big Data und ihre Rolle im Computational Journalism

Zusammenfassung

Der Beitrag strukturiert den aktuellen Forschungsstand zu Computational Journalism. Computer-Assisted Reporting, Data Journalism und Algorithmic Journalism werden als Ausprägungen des Computational Journalism identifiziert und mit Hilfe eines Eingabe-Verarbeitung-Ausgabe-Modells (E-V-A-Modells) voneinander abgegrenzt. Mit diesem Analyseraster kann das Wechselspiel von Mensch, Technik und Daten im Computational Journalism dargestellt werden, indem es sowohl die Rolle digitaler Daten, die Rolle des Algorithmus als auch die Rolle des Journalisten in der Nachrichtenproduktion einbezieht. Eine Diskussion normativer Herausforderungen von Big Data und Algorithmen im Spannungsfeld der Datenbeschaffung, -auswertung und -präsentation sowie ein Ausblick auf zukünftige Forschungsfelder runden den Beitrag ab.

Dörr, K. 2016. Algorithmen, Big Data und ihre Rolle im Computational Journalism. In *Journalismusforschung. Stand und Perspektiven*, herausgegeben von Meier, K., & Neuberger, C., 245–264. Baden-Baden: Nomos.

Copyright: Nomos Verlagsgesellschaft mbH & Co. KG

Genehmigung zum Druck wurde erteilt:

Dr. Sandra Frey

Nomos Verlagsgesellschaft mbH & Co. KG

Lektorat Soziologie, Medien- und Kommunikationswissenschaft

Algorithmen, Big Data und ihre Rolle im Computational Journalism

Konstantin Dörr

„The machines rose from the ashes of the nuclear fire. Their war to exterminate mankind had raged for decades, but the final battle would not be fought in the future. It would be fought here, in our present. Tonight ...“

Terminator (1984)

1 Algorithmus ante portas

Wie der Kampf zwischen Mensch und Maschine am Ende ausgehen wird, ist abseits aller Science-Fiction-Fantasien noch ungewiss. Dennoch stellt sich mit Verweis auf die techniksoziologische Agenten-Forschung und die Forschung zur Künstlichen-Intelligenz¹ (KI) in letzter Instanz die Frage, wie und ob menschliche Leistungen durch Maschinen ersetzt werden können (zur Agenten-Forschung vgl. Harrison/Caglayan 1997; Rammert 2007; zur KI vgl. Minsky 1988; Kurzweil 1990). Weit weniger blutig, aber gleichsam herausfordernd muss sich auch der Journalismus dieser Frage stellen. Es muss geklärt werden, bis zu welchem Grad Technik journalistische Aufgaben bei der Beobachtung und Beschreibung der Gesellschaft übernehmen kann bzw. überhaupt übernehmen sollte. Denn mit der flächendeckenden Einführung des Computers und des Internets in den 1980er- und 1990er-Jahren haben sich die Art und Weise, wie Nachrichten recherchiert, produziert, distribuiert und konsumiert werden, kontinuierlich verändert (vgl. Mitchelstein/Boczkowski 2009; Domingo/Paterson 2011; Powers 2012).

Aktuell steht der Algorithmus im medialen und wissenschaftlichen Rampenlicht (vgl. u.a. Anderson 2012; Gillespie 2014; Napoli 2014; Diakopoulos 2015; Kitchin 2016). Algorithmen regeln, welche Nachrichten

¹ Maschinelle oder auch künstliche Intelligenz wird von Minsky (1988: 326) definiert als „making machines do things that people consider to require intelligence“.

Konstantin Dörr

auf *Twitter* oder im *Facebook*-News-Feed erscheinen. Sie legen fest, welche Suchergebnisse *Google* liefert. Sie empfehlen, welche Nachrichten gerade lesenswert sind, und bestimmen dabei auch, wie sie wahrgenommen werden (vgl. Resnick/Varian 1997; Rieder 2005; Bucher 2012; Anderson 2013; Latzer et al. 2014; Graefe et al. 2016). Algorithmen schätzen die Nachfrage nach journalistischen Produkten ein und beobachten das Publikum. Sie sind Nachrichtenproduzent und Gatekeeper in einer Person (vgl. Dörr 2015; Vos/Heinderyckx 2015). Dabei verschwinden „nicht nur die Grenzen zwischen den Medien, sondern auch die Schnittstellen zwischen den Produktions- und Übermittlungstechniken.“ (Weischenberg/Hienzsch 1994: 478; vgl. Neuberger 2009; Loosen 2015) Zwar ist es Aufgabe des Journalismus, Teilbereiche der Gesellschaft zu beobachten und zu beschreiben, doch wie erste Forschungsergebnisse zeigen, können auch Algorithmen Teile dieser Funktion übernehmen (vgl. Beer 2009; Just/Latzer 2016; Loosen/Scholl 2016). Sie fordern damit direkt den traditionellen Journalismus und seine Akteure heraus. Die ‚Algorithmisierung‘ der Gesellschaft verändert nicht nur die Nachrichtenproduktion, sondern auch Epistemologie, Kompetenzen, Normen und Ökonomie des Journalismus (vgl. Lewis/Westlund 2015b).

Dies zeigt sich besonders im Bereich der computergestützten Recherche, Analyse, Produktion, Empfehlung und Distribution von Inhalten im Bereich des Computational Journalism (vgl. Neuberger 2008; Gynnild 2014; Neuberger/Nuernbergk 2015). Je nach Funktion der Software-Anwendung klassifizieren, strukturieren und filtern Algorithmen elektronische Signale und weisen ihnen Relevanz zu (vgl. Rieder 2004).

Ihr volles Potenzial können diese Anwendungen aber erst im Zusammenspiel mit Big Data entfalten. Unter diesem Begriff werden ökonomische und gesellschaftliche Potenziale und Risiken beim Zugang, der Speicherung und Verarbeitung großer, unstrukturierter digitaler Daten diskutiert (vgl. Flew et al. 2012; Mayer-Schönberger/Cukier 2013). Big Data ist dabei keine neue Technologie, sondern vielmehr eine neue Methode der Erkenntnisgewinnung, mit der durch algorithmische Prozesse Komplexität reduziert wird (vgl. Lewis 2015; Mayer-Schönberger 2015).

Dieser Beitrag beleuchtet sowohl die Rolle digitaler Daten, die Rolle des Algorithmus wie auch die Rolle des Journalisten für die Nachrichtenproduktion im Computational Journalism. Darauf aufbauend wird – eng geknüpft an die Funktionsweise des Computers – dieses Verhältnis abstrakt anhand eines Eingabe-Verarbeitung-Ausgabe-Modells (EVA-Mo-

dell) dargestellt.² Die Diskussion normativer Herausforderungen von Big Data und Algorithmen bei der Datenbeschaffung, Auswertung und Präsentation sowie der Ausblick auf zukünftige Forschungsfelder runden den Beitrag ab.

2 Der Computational Journalism und seine Varianten

Vor dem Hintergrund der Digitalisierung und gefördert durch Big Data hat sich mittlerweile der Begriff ‚Computational Journalism‘ als Bezeichnung für eine Vielzahl technischer Innovationen im Journalismus durchgesetzt (vgl. Neuberger/Nuernbergk 2015). Computer-Assisted-Reporting, Data Journalism und Algorithmic Journalism können darunter subsumiert werden (vgl. Hamilton/Turner 2009; Anderson 2012; ähnlich Berger et al. 2015; für einen weiteren Systematisierungsvorschlag vgl. Coddington 2015). Doch diese Differenzierungen sind durch die Vielzahl an praxisgetriebenen Definitionen eine Herausforderung. Mit der technischen Entwicklung erweitern sich einerseits die journalistischen Einsatzmöglichkeiten, andererseits fällt es dadurch auch immer schwerer, die neuen Anwendungen bestehenden Definitionen und Begriffen zuzuordnen. Denn verschiedene Eigenschaften des Computational Journalism, z.B. die Nutzung des Computers oder digitaler Daten, treffen gleich auf mehrere Entwicklungsformen zu. Die unterschiedliche Verwendung zentraler Begriffe aus der englischsprachigen Forschung zeigt sich auch bei Übersetzungsversuchen ins Deutsche. Um Unklarheiten zu vermeiden, stützt sich dieser Beitrag daher maßgeblich auf die englischsprachigen Begriffe.

Flew und Kollegen (2012: 157) fassen unter Computational Journalism die „application of computing to journalism“. Diese Charakterisierung umfasst beides: den Einsatz computergestützter Werkzeuge und ihre Verbindung mit quantitativen Methoden der Analyse von Big Data (vgl. Hamilton/Turner 2009). Im Kontext dieses „quantitative turn“ (Lewis/Westlund 2015b: 452) verweist Diakopoulos (2011) auch auf die unterschiedlichen Phasen des Computational Journalism: auf die Informationssammlung, Organisation und Sinngebung, Kommunikation und Präsentation, Verbreitung und Resonanz von Big Data in der journalistischen Praxis. Im

2 Ein Computer verarbeitet Eingabedaten in Ausgabedaten, ein grundsätzliches Prinzip – das EVA-Prinzip (Eingabe – Verarbeitung – Ausgabe) ist auch auf Software bzw. Algorithmen übertragbar (vgl. Latzer et al. 2014).

Konstantin Dörr

Folgenden werden seine Ausprägungen mit Verweis auf die Rolle des Journalisten, des Algorithmus und Big Data in der Nachrichtenproduktion differenziert.

2.1 Computer-Assisted-Reporting (CAR)

Der technische Fortschritt hat dazu geführt, dass der Begriff ‚Computer-Assisted-Reporting‘ (CAR) mittlerweile breiter verwendet wird. Die Anfänge des CAR liegen bereits in den 1950er-Jahren, als der US-amerikanische Fernsehsender *CBS* Daten zur Präsidentenwahl mit der Hilfe von Computern analysierte. Unter CAR versteht man grundsätzlich: „[U]sing a computer to help report a story“ (Mayo/Leshner 2000: 71; vgl. Cox 2000; Meyer 2002; Flew et al. 2012). Erstmals wurden auch sozialwissenschaftliche Verfahren der Datenanalyse genutzt, die den investigativen Journalismus förderten (vgl. Coddington 2015). Dennoch handelte es sich bei diesen Techniken zunächst um Randphänomene, auch weil der Einzug des Computers in die Redaktionen erst allmählich in den 1980er-Jahren stattfand und die Verbreitung des Internets der Entwicklung erst in den 1990er-Jahren einen Schub verlieh (vgl. Neuberger 2009). Diese Entwicklung fällt zusammen mit ersten Fortschritten in der KI-Forschung: Ende der 1960er-Jahre entwickelte Joseph Weizenbaum das Programm *ELIZA*, in dem der Dialog eines Psychotherapeuten mit einem Patienten simuliert wurde. 1997 besiegte das System *Deep Blue* von *IBM* den Schach-Weltmeister Garri Kasparov in der ersten von insgesamt sechs Partien. Das erste Mal in der Geschichte das Verhältnisses ‚Mensch gegen Maschine‘ gewann der Computer. Im Journalismus war der Computer in diesem Entwicklungsstadium aber zunächst nur ein einfaches Werkzeug des Journalisten.

Auch wenn die Vielfalt der Anwendungen im Computer-Assisted-Reporting eine eindeutige Funktionalitätsbeschreibung erschwert, so zeigt sich dennoch ein eindeutiges Muster: Algorithmen agieren nur in klar definierten Strukturen und nach bestimmten Regeln und führen dabei die Anfrage des Journalisten aus (vgl. Tab. 1). Der Journalist kontextualisiert die Inhalte – er schreibt den Artikel –, und entscheidet über deren Publikation und Distribution. Auch die Verfügbarkeit und Möglichkeiten zur Speicherung digitaler Daten waren in diesem Stadium sehr eingeschränkt. Der *Apple Lisa*, einer der ersten Personal-Computer, hatte bei seiner Markteinführung im Jahr 1983 einen externen Festplattenspeicher von 10 MB.

2.2 Data Journalism

Der Data Journalism³ steht in der Tradition des Computer-Assisted-Reporting. Doch mit der technischen Entwicklung und der steigenden Verfügbarkeit von Big Data ergibt sich die Notwendigkeit einer weiteren Differenzierung. Diese Form des Computational Journalism betont den Datenaspekt und befasst sich thematisch mit strukturellen Herausforderungen, die sich quantitativ beschreiben und erklären lassen (vgl. Stray 2011; Neuberger/Nuernbergk 2015). Das Konzept umfasst sowohl die explorative wie auch die zielgerichtete journalistische Recherche in großen, digitalen und öffentlich verfügbaren Datensätzen (Big Data, Open Data), die Verarbeitung der ausgewählten Daten mithilfe algorithmischer Prozesse als auch die Darstellung in Form von Auswertungen, Tabellen und Visualisierungen. Der Fokus des Data Journalism liegt auf digitalem Storytelling (vgl. u.a. Bell 2012; Weber/Rall 2013). Der Journalist erzählt mithilfe der Daten eine Geschichte, bindet sie in einen redaktionellen Kontext ein und produziert aus diesen Daten einen Mehrwert. Datenquellen sind z.B. WikiLeaks und OffshoreLeaks⁴. Journalist und Publikum – nicht der Algorithmus – reflektieren und kontextualisieren dabei die Formen der Datenverwendung und beschreiben so gesellschaftliche Veränderungen (vgl. Langer 2011; Gray/Bounegru/Chambers 2012; Coddington 2015). Algorithmen sind ebenfalls nur Werkzeug und führen die Eingaben des Journalisten nach bestimmten Regeln aus. Neben der Voraussetzung des Datenzugangs⁵ kommt es im Data Journalism auch auf individuelle Fähigkeiten des Journalisten an (z.B. Programmier- und Statistikkenntnisse). Zum einen, um die verfügbaren Daten auszuwerten, und zum anderen, um diese auch entsprechend interpretieren zu können (vgl. Yarnall et al. 2008). Doch Data Journalism ist aufwendig und kostenintensiv. Nicht zuletzt ist dies auch ein Grund, warum diese Form des Computational Journalism oft noch in kleinen Teams, bestehend aus Programmierern und Journalisten, und eher punktuell für einzelne Themen oder Ereignisse stattfindet (vgl. Weinacht/Spiller 2014).

3 ‚Big Data Journalism‘ oder ‚Data-Driven Journalism‘ sind Begriffe, die in diesem Kontext oftmals synonym verwendet werden.

4 Eine Recherchedatenbank des International Consortium of Investigative Journalists: <https://offshoreleaks.icij.org/>.

5 Die öffentliche Verwaltung in Großbritannien stellt beispielsweise Daten zu ihren Tätigkeiten auf <http://data.gov.uk> zur Verfügung.

Konstantin Dörr

2.3 Algorithmic Journalism

Unter den Begriffen „Robot Journalism“ (Clerwall 2014), „Automated Journalism“ (Carlson 2015; Graefe 2016) oder „Algorithmic Journalism“ (van Dalen 2012; Dörr 2015) werden aktuell u.a. technische, ökonomische, rechtliche und ethische Herausforderungen der automatisierten Texterstellung diskutiert (vgl. Weeks 2014; Young/Hermida 2015; Carlson 2015; Dörr/Hollnbuchner 2016). Wie im Fall des Quakebot für Erdbebennachrichten der *Los Angeles Times* oder der automatisierten Geschäftsberichte der *Associated Press* binden traditionelle journalistische Akteure diese Inhalte bereits fest in ihr Medienangebot ein. Auch in Deutschland experimentieren Tageszeitungen wie die *Berliner Morgenpost* mit dem Feinstaubmonitor mit solchen Software-Anwendungen. Voraussetzung für diese Art der Automatisierung ist die Verfügbarkeit von strukturierten, digitalen Daten. Auf dieser Grundlage können Software-Anbieter wie *Automated Insights* mit ihrer Plattform Wordsmith automatisierte Texte erstellen. Die Formung von Meinungen und die Einordnung von Ereignissen in gesellschaftliche Gesamtzusammenhänge kann diese Form des Journalismus zwar nicht leisten. Es geht hier jedoch weniger um die Substitution des Journalisten und seiner kreativen und intellektuellen Leistungen. Vielmehr ermöglichen die Technologie und die Verfügbarkeit von Big Data die massenhafte, standardisierte Textproduktion für spezielle Themenbereiche wie Sport, Wetter oder Finanzen.

Ein großer Teil der Forschung konzentriert sich auf das Rezeptionsverhalten und speziell auf Fragen zur Wahrnehmung und Glaubwürdigkeit dieser Texte (vgl. Clerwall 2014; van der Kaa/Krahmer 2014; Graefe et al. 2016). Es zeigt sich u.a., dass Leser beim Vergleich kurzer Texte fast keine Unterschiede zwischen menschlich und algorithmisch geschriebenen Versionen erkennen. Diese automatisierte Produktion – von der Auswahl über die Verarbeitung bis zur Darstellung – wirft Fragen zur Rolle von Big Data, der Journalisten und Algorithmen in diesem Prozess auf.

Automatisierungsprozesse waren und sind seit jeher Teil der technischen Evolution im Journalismus, z.B. die Entwicklung vom Zeitungsdruck zum Digital Publishing (vgl. Stöber 2008). Auch Algorithmic Journalism ist nicht als ein isoliertes Phänomen zu betrachten, sondern in diese Entwicklung einzuordnen. Doch im Vergleich zu CAR oder Data Journalism zeigt sich ein zentraler Unterschied. Mit Verweis auf das EVA-Modell operieren auf der Ebene der Eingabe Algorithmen zwar ebenfalls nach festgelegten Routinen und führen die Eingaben des Journalisten aus. Inter-

essanter ist jedoch die Ebene der Verarbeitung. In Bezug auf Sprache, aber auch bei der Kontextualisierung und Beschreibung von Inhalten können Algorithmen hier unabhängig von menschlicher Einflussnahme auch anders handeln und eine Eigendynamik entwickeln. Während der Journalist keinen unmittelbaren Einfluss auf die Ebene der Verarbeitung hat, zeigen sich gewisse Freiheitsgrade in der Art und Weise, wie die Eingabeaufforderung des Journalisten ausgeführt wird. Je nachdem, wie vielfältig das zugrundeliegende sprachliche Regelwerk ausfällt, kann der Algorithmus auf dieser Ebene beispielsweise Synonyme für Wortgruppen erstellen (z.B. schreibt er statt ‚groß‘ ‚riesig‘ oder ‚gewaltig‘), beim Satzanfang variieren, andere Adjektive verwenden oder einzelne inhaltliche Aspekte hervorheben. Im Rahmen dieser menschlich erzeugten Regelwerke kann die Technik hier autonom entscheiden (vgl. Dörr 2015). Der Grad der Automatisierung innerhalb der Varianten des Computational Journalism wird mit der technischen Entwicklung und der steigenden Verfügbarkeit von Big Data zunehmen. Diese Prozesse reichen dabei von der einfachen Nutzung computerisierter Prozesse als journalistisches Werkzeug über komplexere Anwendungen des Data Journalism bis hin zu ersten Formen der KI im Algorithmic Journalism. Das Verhältnis zwischen Journalist, Algorithmus und Big Data ist zusammenfassend in Anlehnung an das EVA-Modell in Tabelle 1 dargestellt.

Tabelle 1: Die Rolle von Big Data, Algorithmen und Journalisten für die Nachrichtenproduktion im Computational Journalism

252

	Eingabe	Verarbeitung	Ausgabe	Abgrenzung
Computer-Assisted-Reporting	<u>Rolle des Journalisten:</u> - Festlegung des Zwecks und Ziel der Anwendung - Wahl des Datensatzes <u>Rolle digitaler Daten:</u> - privat/öffentlich verfügbar - Basis für zielgerichtete Anwendung, z.B. Recherche, Datenanalyse	<u>Rolle des Journalisten:</u> - Kontrolle <u>Rolle der Algorithmen:</u> - Selektion und Relevanzzuweisung auf Basis vorher festgelegter Regeln - Kein Handlungsspielraum/Freiheitsgrade	<u>Rolle des Journalisten:</u> - Kontrolle - Kontextualisierung und Interpretation der Inhalte - Publikation <u>Rolle der Algorithmen:</u> - Ausführung der Eingabe - Darstellung der Ergebnisse nach festgelegter Programmierung der jeweiligen Anwendung	- Computer und Algorithmen als Werkzeug - Frühe Phase der KI-Forschung - Anwendungen zur Recherche, Analyse - Begrenzte Möglichkeiten beim Zugang, der Speicherung und Verarbeitung digitaler Daten - Fokus und Förderung des investigativen Journalismus
Data Journalism	<u>Rolle des Journalisten:</u> - Festlegung des Zwecks und Ziel der Anwendung - Wahl des Datensatzes <u>Rolle digitaler Daten:</u> - Big Data - privat/öffentlich verfügbar - Basis für zielgerichtete Anwendung, z.B. Visualisierung	<u>Rolle des Journalisten:</u> - Kontrolle <u>Rolle der Algorithmen:</u> - Selektion und Relevanzzuweisung auf Basis vorher festgelegter Regeln - Kein Handlungsspielraum/Freiheitsgrade	<u>Rolle des Journalisten:</u> - Kontrolle - Kontextualisierung und Interpretation der Inhalte - Publikation <u>Rolle der Algorithmen:</u> - Ausführung der Eingabe - Darstellung der Ergebnisse nach festgelegter Programmierung der jeweiligen Anwendung	- Ausdifferenzierung von Computer-Assisted-Reporting - Computer und Algorithmen als Werkzeug - Entwicklung geprägt durch Verfügbarkeit von Big Data und technischem Fortschritt - Anwendungen zur Recherche, Analyse oder Visualisierung - Fokus auf Digital Storytelling

Algorithmen, Big Data und ihre Rolle im Computational Journalism

Eingabe	Verarbeitung	Ausgabe	Abgrenzung
<p><u>Rolle des Journalisten:</u></p> <ul style="list-style-type: none"> - Festlegung des Zwecks und Ziel der Anwendung - Wahl des Datensatzes - Bestimmung allg. Parameter, z.B. Ort, Zeit, Länge des Textes <p><u>Rolle digitaler Daten:</u></p> <ul style="list-style-type: none"> - Big Data - privat/öffentlich verfügbar - Basis für Textgenerierung 	<p><u>Rolle des Journalisten:</u></p> <ul style="list-style-type: none"> - passiv, kein Einfluss <p><u>Rolle der Algorithmen:</u></p> <ul style="list-style-type: none"> - Selektion und Relevanzzuweisung auf Basis vorher festgelegter Regeln - Handlungsspielraum/Freiheitsgrade bei der Wahl und Gewichtung der Daten 	<p><u>Rolle des Journalisten:</u></p> <ul style="list-style-type: none"> - Kontrolle (aufgrund der Vielzahl und Schnelligkeit der Generierung aber nur eingeschränkt möglich, ex post) <p><u>Rolle der Algorithmen:</u></p> <ul style="list-style-type: none"> - Darstellung und Publikation nach festgelegter Programmierung der jeweiligen Anwendung - Freiheitsgrade bei der Darstellung/Kontextualisierung möglich - Publikation nach initialer Programmierung vollautomatisiert - Varianz beim Zeitpunkt der Publikation möglich 	<ul style="list-style-type: none"> - Ausdifferenzierung von Data Journalism - Fortschritte in der KI-Forschung - Computer und Algorithmen zwischen Werkzeug und eigenständigem Akteur - Anwendungen zur automatisierten Textproduktion (u.a. Sport, Finanzen, Wetter) - Big Data als Grundlage

Algorithmic Journalism

Konstantin Dörr

3 Normative Herausforderungen für den Computational Journalism

Vor dem Hintergrund dieser Ausdifferenzierung und der Veränderung journalistischer Aufgaben- und Tätigkeitsfelder ergeben sich zahlreiche normative Herausforderungen. Denn mit der steigenden Komplexität der Interaktion zwischen Journalist und Algorithmus und den damit verbundenen Automatisierungsprozessen erhöhen sich die Anforderungen an professionelle Akteure auf individueller und organisatorischer Ebene (vgl. Köberer 2014). Die journalistische Ethik kann für die Diskussion dieser Fragestellungen hilfreich sein (vgl. Debatin 1998). Sie versucht, normativ journalistisches Handeln zu begründen und Orientierung für die journalistische Praxis zu stiften.

Mit Blick auf das journalistische Handlungsfeld und der Frage, welche normativen Herausforderungen insbesondere für den Computational Journalism von Bedeutung sind, lassen sich Problemfelder auf der Ebene der Datenbeschaffung, der algorithmisch-gesteuerten Datenauswertung und der Präsentation erkennen.

In der Wissenschaft werden die Herausforderungen von Big Data und Algorithmen seit einiger Zeit kritisch reflektiert (vgl. Boyd/Crawford 2013; Ananny 2016). Sie zeigen sich bei der Datenauswahl, Datensuche und Herkunft (vgl. Zion/Craig 2014), der Objektivität (vgl. Gillespie 2014), der Transparenz (vgl. McBride/Rosenstiel 2013) und bei der Frage, wie journalistische Werte im Programmcode berücksichtigt werden können (vgl. Kraemer/van Overveld/Peterson 2011; Young/Hermida 2015). Die Verfügbarkeit und der Zugang zu digitalen Daten sind dabei insbesondere für den Data Journalism und den Algorithmic Journalism essentiell. Die Rolle des Journalisten und die korrekte Kontextualisierung dieser Daten wird immer wichtiger. Fragen nach der Herkunft der Daten, deren Objektivität und Repräsentativität sind genauso relevant wie Fragen der Verantwortungsübernahme und Kontrolle. Dabei sind ethische Kodizes für den Umgang mit Big Data und Algorithmen im Computational Journalism noch Mangelware (vgl. Pasquale 2015). Stattdessen wird konsequent auf das Gebot der Transparenz verwiesen (vgl. u.a. Diakopoulos 2015; Dörr/Hollnbuchner 2016). Diakopoulos (2015) nennt verschiedene Transparenzregeln für die algorithmische Berichterstattung: Es soll kommuniziert werden, welche Auswahlkriterien verwendet werden, welche Datenquellen genutzt werden, ein möglicher Bias soll durch die Angabe fehlerhafter Datenzuordnungen vermieden werden, zudem sollen Bewertungsmaßstäbe für die Klassifikation der Daten entwickelt werden. Korrekturen müssen

ebenfalls im Sinne der Transparenz kenntlich gemacht werden. Wie am Beispiel einer Nachrichtenmeldung auf *Fox News* gezeigt werden kann, kann Algorithmic Journalism auf Basis mangelnder Daten oder ungenauer Programmierung auch Fehler produzieren. Diese müssen – wie hier geschehen – im Sinn des Transparenzgebots kenntlich gemacht werden:

„In a story Aug. 5 about Graham Holdings Co. earnings — generated by Automated Insights using data from Zacks Investment Research — The Associated Press reported incorrectly that Graham Holdings shares had fallen since the start of the year and a year ago.”

Doch dabei stellt sich insbesondere die Frage, wie diese Regeln auch außerhalb professionell-journalistischer Strukturen umgesetzt werden können. Es ist davon auszugehen, dass nicht-journalistisch sozialisierte Akteure wie Programmierer, Software- bzw. Daten-Anbieter oder auch Plattformbetreiber wie *Facebook* eine immer wichtigere Rolle in der Nachrichtenproduktion und -distribution einnehmen werden. Denn mit der steigenden Komplexität der Anwendungen und der algorithmischen Prozesse ergeben sich Rollenverschiebungen und -erweiterungen zwischen Journalisten, diesen externen Akteuren, Technik und Publikum (vgl. Anderson 2011). Dies trifft insbesondere bei der Datenauswertung und Präsentation zu. Während im Data Journalism Algorithmen strikt nach der Programmierung und der Eingabe der Journalisten die Arbeitsschritte ausführen, haben sie im Algorithmic Journalism gewisse Freiheitsgrade bei der Texterstellung und Publikation. Aufgrund der steigenden Komplexität der Anwendungen ist unklar, welche Selektions- und Relevanzzuweisungsprozesse im Hintergrund ablaufen. Dieses Black-Box-Problem gilt es zu lösen, will man die Herausforderungen für den Journalismus unter normativen Gesichtspunkten analysieren (vgl. Diakopoulos 2015; Lewis 2015).

4 Fazit und Ausblick

Das Verhältnis von Mensch und Maschine im Computational Journalism und der Einsatz von Big Data entwickeln sich mit dem technischen Fortschritt und der Verknüpfung mit Ansätzen der KI-Forschung stetig weiter. Elemente des Machine Learning findet man bereits im Algorithmic Journalism. Darunter versteht man das automatisierte Lernen von Algorithmen anhand von Mustern (z.B. für aktuelle Wahlprognosen für die US-Wahl 2016: <http://pollyvote.com>). Kombiniert wird diese Form der Mustererkennung oft mit Elementen des Data Mining, womit die Suche nach Mus-

Konstantin Dörr

tern in Big Data gemeint ist. Das Internet ist dabei nicht nur Medium zur Speicherung und Distribution von Inhalten, sondern auch ein Netz aus Computern (vgl. Neuberger 2009).

Wie die Formen des Computational Journalism zeigen, können die Computer dabei digitale Daten unterschiedlich bearbeiten. Sie regieren dabei nicht nur auf Anweisung des Journalisten, sondern haben auch die Fähigkeit, autonom und lernfähig zu sein.

Nicht zuletzt tragen diese Prozesse zu einer weiteren Entgrenzung des Journalismus bei und verändern die Anforderungen an die professionellen Akteure. Sie müssen die zentralen Leistungen des Journalismus bei der Beobachtung und Beschreibung der Gesellschaft auch in diesem Spannungsfeld erbringen. Rollenverschiebungen und -erweiterungen lassen sich nicht nur zwischen Journalist und Algorithmus diskutieren, auch der Einfluss von journalismusfernen Akteuren und dem Publikum muss im Hinblick auf die Nachrichtenproduktion, Distribution und Rezeption untersucht werden. Ökonomische Interessen von Software-Anbietern, Programmierern und Plattformbetreibern beeinflussen sowohl die inhaltliche Gestaltung des Journalismus als auch die Redaktionsorganisation. Auch das Publikum entscheidet mit seinem Nutzungsverhalten darüber, wie sich die Ausprägungen des Computational Journalism weiterentwickeln.

Bei normativen Fragen zur Transparenz und Verantwortung können medienethische Konzepte hilfreich sein. Hier gilt es insbesondere zu klären, ob Algorithmen als Träger von Verantwortung gelten können. Dies ist insofern relevant, als ethische Fragestellungen immer subjektgebunden sind und rationales Handeln bisher nur dem Menschen zugeschrieben wurde.

Mit der Entwicklung der KI und der Möglichkeit der Bewertung von Inhalten durch selbstständige Algorithmen können auch techniksoziologische Fragestellungen, z.B. nach einer Handlungsträgerschaft von Technik, einbezogen werden (vgl. Rammert/Schulz-Schaeffer 2002; Neff et al. 2012; Lewis/Westlund 2015a). Schreibt man Technik Handlungsfähigkeit zu, dann tritt sie in Wechselbeziehung zu menschlichen Akteuren und wird damit Teil gemeinschaftlicher Tätigkeiten (vgl. Matsuzaki 2011). Von der Offenheit gegenüber solchen Ansätzen könnte die Journalismusforschung profitieren; dafür müsste sie ihren bislang auf Menschen oder Organisationen bezogenen Akteursbegriff erweitern.

Literatur

- Ananny, Mike (2016): Toward an Ethics of Algorithms: Convening, Observation, Probability, and Timeliness. In: *Science, Technology, & Human Values*, 41. Jg., H. 1, S. 93–117.
- Anderson, Chris W. (2011): Deliberative, Agonistic, and Algorithmic Audiences: Journalism's Vision of Its Public in an Age of Audience Transparency. In: *International Journal of Communication*, 5. Jg., S. 529–547.
- Anderson, Chris W. (2012): Towards a Sociology of Computational and Algorithmic Journalism. In: *new media & society*, 15. Jg., H. 7, S. 1005–1021.
- Anderson, Chris W. (2013): What aggregators do: Towards a networked concept of journalistic expertise in the digital age. In: *Journalism*, 14. Jg., H. 8, S. 1008–1023.
- Beer, David (2009): Power through the algorithm? Participatory web cultures and the technological unconscious. In: *new media & society*, 11. Jg., H. 6, S. 985–1002.
- Bell, Emily (2012): Journalism by numbers. In: *cjr.org*. http://www.cjr.org/cover_story/journalism_by_numbers.php?page=all. Zugegriffen am 16. Juni 2016.
- Berger, Matthias/Haim, Mario/Graefe, Andreas/Brosius, Hans-Bernd/Hess, Thomas (2015): Computational Journalism. In: *MedienWirtschaft*, 12. Jg., H. 1, S. 22–25.
- Boyd, Danah/Crawford, Kate (2013): Big Data als kulturelles, technologisches und wissenschaftliches Phänomen. Sechs Provokationen In: Heinrich Geiselberger/Tobias Moorstedt (Hg.): *Big Data. Das neue Versprechen der Allwissenheit*. Berlin: Suhrkamp, S. 187–218.
- Bucher, Taina (2012): Want to be on the top? Algorithmic power and the threat of invisibility on Facebook. In: *new media & society*, 14. Jg., H. 7, S. 1164–1180.
- Carlson, Matt (2015): The Robotic Reporter. In: *Digital Journalism*, 3. Jg., H. 3, S. 416–431.
- Clerwall, Christer (2014): Enter the Robot Journalist. Users' perceptions of automated content. In: *Journalism Practice*, 8. Jg., H. 5, S. 519–531.
- Coddington, Mark (2015): Clarifying Journalism's Quantitative Turn. A typology for evaluating data journalism, computational journalism, and computer-assisted reporting. In: *Digital Journalism*, 3. Jg., H. 3, S. 331–348.
- Cox, Melisma (2000): The Development of Computer-assisted Reporting. In: *Association for Education in Journalism and Mass Communication*, 17–18 März, University of North Carolina. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.631.6220&rep=rep1&type=pdf>. Zugegriffen am 16. Mai 2016.
- Debatin, Bernhard (1998): Verantwortung im Medienhandeln. Medienethische und handlungstheoretische Überlegungen zum Verhältnis von Freiheit und Verantwortung in der Massenkommunikation. In: Wolfgang Wunden (Hg.): *Freiheit und Medien*. Frankfurt a.M.: Abt. Verlag, S. 113–130.
- Diakopoulos, Nicholas (2015): Algorithmic Accountability. In: *Digital Journalism*, 3. Jg., H. 3, S. 398–415.
- Diakopoulos, Nick (2011): A functional roadmap for computational journalism. In: *nickdiakopoulos.com*. http://www.nickdiakopoulos.com/wp-content/uploads/2007/05/CJ_Whitepaper_Diakopoulos.pdf. Zugegriffen am 21. April 2016.

Konstantin Dörr

- Domingo, David/Paterson, Chris (Hg.) (2011): *Making Online News: Newsroom Ethnographies in the Second Decade of Internet Journalism* (2. Aufl.). New York: Peter Lang.
- Dörr, Konstantin (2015): Mapping the Field of Algorithmic Journalism. In: *Digital Journalism* (online first), S. 1–24 (<http://dx.doi.org/10.1080/21670811.2015.1096748>).
- Dörr, Konstantin/Hollnbuchner, Katharina (2016): Ethical Challenges of Algorithmic Journalism. In: *Digital Journalism* (online first), S. 1–16 (<http://dx.doi.org/10.1080/21670811.2016.1167612>).
- Flew, Terry/Spurgeon, Christina/Daniel, Anna/Swift, Adam (2012): The promise of computational Journalism. In: *Journalism Practice*, 6. Jg., H. 2, S. 157–171.
- Gillespie, Tarleton (2014): The relevance of algorithms. In: Tarleton Gillespie/Pablo J. Boczkowski/Kirsten A. Foot (Hg.): *Media technologies. Essays on communication, materiality, and society*. Cambridge, MA: MIT Press, S. 167–193.
- Graefe, Andreas (2016): Guide to Automated Journalism. In: [towcenter.gitbooks.io](https://towcenter.gitbooks.io/guide-to-automated-journalism/content/). <https://towcenter.gitbooks.io/guide-to-automated-journalism/content/>. Zugriffen am 12. April.
- Graefe, Andreas/Haim, Mario/Haarmann, Bastian/Brosius, Hans-Bernd (2016): Readers' perception of computer-generated news: Credibility, expertise, and readability. In: *Journalism* (online first), S. 1–16 (<http://dx.doi.org/10.1177/1464884916641269>).
- Gray, Jonathan/Bounegru, Liliana/Chambers, Lucy (2012): *The Data Journalism Handbook – How Journalists Can Use Data to Improve the News*. Sebastopol: O'Reilly.
- Gynnild, Astrid (2014): Journalism innovation leads to innovation journalism: The impact of computational exploration on changing mindsets. In: *Journalism*, 15. Jg., H. 6, S. 713–730.
- Hamilton, James T./Turner, Fred (2009): Accountability Through Algorithm. Developing the Field of Computational Journalism. A report from Developing the Field of Computational Journalism, Center For Advanced Study in the Behavioral Sciences Summer Workshop, July 27–31. In: [web.stanford.edu](http://web.stanford.edu/~fturner/Hamilton%20Turner%20Acc%20by%20Alg%20Final.pdf). <http://web.stanford.edu/~fturner/Hamilton%20Turner%20Acc%20by%20Alg%20Final.pdf>. Zugriffen am 30. April 2016.
- Harrison, Colin/Caglayan, Alper (1997): *Agent Sourcebook: A Complete Guide to Desktop, Internet, and Intranet Agents*. New York: John Wiley & Sons.
- Just, Natascha/Latzer, Michael (2016): Governance by algorithms: reality construction by algorithmic selection on the Internet. In: *Media, Culture & Society* (online first), S. 1–21 (<http://dx.doi.org/10.1177/0163443716643157>).
- Kitchin, Rob (2016): Thinking critically about and researching algorithms. In: *Information, Communication & Society* (online first), S. 1–17 (<http://dx.doi.org/10.1080/1369118X.2016.1154087>).
- Köberer, Nina (2014): *Advertorials in Jugendprintmedien. Ein medienethischer Zugang*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Kraemer, Felicitas/van Overveld, Klees/Peterson, Martin (2011): Is There an Ethics of Algorithms? In: *Ethics and Information Technology*, 13 Jg., H. 3, S. 251–260.

Algorithmen, Big Data und ihre Rolle im Computational Journalism

- Kurzweil, Ray (1990): *The Age of Intelligent Machines*. Cambridge MA: MIT Press.
- Langer, Ulrike (2011): *Datenjournalismus*. Journalisten Werkstatt. Salzburg-Eugendorf: Medienfachverlag Oberbauer.
- Latzer, Michael/Hollnbuchner, Katharina/Just, Natascha/Saurwein, Florian (2014): The economics of algorithmic selection on the Internet. Working Paper, University of Zurich, Zürich. In: *Mediachange.ch*. http://www.mediachange.ch/media/pdf/publications/Economics_of_algorithmic_selection_WP.pdf. Zugegriffen am 20. April 2016.
- Lewis, Seth C. (2015): Journalism In An Era Of Big Data. In: *Digital Journalism*, 3. Jg., H. 3, S. 321–330.
- Lewis, Seth C./Westlund, Oscar (2015a): Actors, Actants, Audiences, and Activities in Cross-Media News Work. In: *Digital Journalism*, 3. Jg., H. 1, S. 19–37.
- Lewis, Seth C./Westlund, Oscar (2015b): Big Data and Journalism. Epistemology, Expertise, Economics, and Ethics. In: *Digital Journalism*, 3. Jg., H. 3, S. 447–466.
- Loosen, Wiebke (2015): The Notion of the “Blurring Boundaries”. In: *Digital Journalism*, 3. Jg., H. 1, S. 68–84.
- Loosen, Wiebke/Scholl, Armin (2016): Journalismus im Zeitalter algorithmischer Wirklichkeitskonstruktion. Vortrag auf der 61. Jahrestagung der DGPK, Leipzig.
- Matsuzaki, Hironori (2011): Die Frage nach der 'Agency' von Technik und die Normenvergessenheit der Techniksoziologie. In: Nico Lüdtk/Hironori Matsuzaki (Hg.): *Akteur – Individuum – Subjekt: Fragen zu 'Personalität' und 'Sozialität'*. Wiesbaden: VS Verlag für Sozialwissenschaften, S. 301–325.
- Mayer-Schönberger, Viktor (2015): Was ist Big Data? Zur Beschleunigung des menschlichen Erkenntnisprozesses. In: *bpb.de*. <https://www.bpb.de/apuz/202242/zur-beschleunigung-menschlicher-erkenntnis?p=all>. Veröffentlicht am 6. März 2015; zugegriffen am 20. April 2016.
- Mayer-Schönberger, Viktor/Cukier, Kenneth (2013): *Big Data. Die Revolution, die unser Leben verändern wird*. München: Redline.
- Mayo, Justin/Leshner, Glenn (2000): Assessing the credibility of computer-assisted reporting. In: *Newspaper Research Journal*, 21. Jg., H. 4, S. 68–82.
- McBride, Kelly/Rosenstiel, Tom (2014): *The New Ethics of Journalism*. Thousand Oaks, CA: CQ Press.
- Meyer, Philip (2002): *Precision Journalism. A Reporter's Introduction to Social Science Methods*. Lanham: Rowman & Littlefield.
- Minsky, Marvin (1988): *The Society of Mind*. New York: Simon and Schuster.
- Mitchelstein, Eugenia/Boczkowski, Pablo J. (2009): Between Tradition and Change: A Review of Recent Research on Online News Production. In: *Journalism*, 10. Jg., H. 5, S. 562–586.
- Napoli, Philip M. (2014): Automated Media: An Institutional Theory Perspective on Algorithmic Media Production and Consumption. In: *Communication Theory*, 24. Jg., H. 3, S. 340–360.

Konstantin Dörr

- Neff, Gina/Jordan, Tim/McVeigh-Schultz, Joshua/Gillespie, Tarleton (2012): Affordances, Technical Agency, and the Politics of Technologies of Cultural Production. In: *Journal of Broadcasting & Electronic Media*, 56. Jg., H. 2, S. 299–313.
- Neuberger, Christoph (2008): Neue Medien als Herausforderung für die Journalismustheorie: Paradigmenwechsel in der Vermittlung öffentlicher Kommunikation. In: Carsten Winter/Andreas Hepp/Friedrich Krotz (Hg.): *Theorien der Kommunikations- und Medienwissenschaft. Grundlegende Diskussionen, Forschungsfelder und Theorieentwicklungen*. Wiesbaden: VS Verlag für Sozialwissenschaften, S. 251–267.
- Neuberger, Christoph (2009): Internet, Journalismus und Öffentlichkeit. Analyse des Medienumbruchs. In: Christoph Neuberger/Christian Nuernbergk/Melanie Rischke (Hg.): *Journalismus im Internet: Profession – Partizipation – Technisierung*. Wiesbaden: VS Verlag für Sozialwissenschaften, S. 19–105.
- Neuberger, Christoph/Nuernbergk, Christian (2015): Verdatete Selbstbeschreibung der Gesellschaft – Über den Umgang des Journalismus mit Big Data und Algorithmen. In: Florian Süssenguth (Hg.): *Die Gesellschaft der Daten – Über die digitale Transformation der sozialen Ordnung*. Bielefeld: transcript, S. 165–189.
- Pasquale, Frank (2015): *The Black Box Society*. Cambridge: Harvard University Press.
- Powers, Matthew (2012): 'In Forms That Are Familiar and Yet-to-be Invented': American Journalism and the Discourse of Technologically Specific Work. In: *Journal of Communication Inquiry*, 36. Jg., H. 1, S. 24–43.
- Rammert, Werner (2007): *Technik – Handeln – Wissen. Zu einer pragmatistischen Technik- und Sozialtheorie*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Rammert, Werner/Schulz-Schaeffer, Ingo (2002): Technik und Handeln. Wenn soziales Handeln sich auf menschliches Handeln und technische Abläufe verteilt. In: Werner Rammert/Ingo Schulz-Schaeffer (Hg.): *Können Maschinen handeln? Soziologische Beiträge zum Verhältnis von Mensch und Technik*. Frankfurt a.M.: Campus, S. 11–64.
- Resnick, Paul/Varian, Hal R. (1997): Recommender systems (Special Section: Recommender Systems). In: *Communications of the ACM*, 40. Jg., H. 3, S. 56–58.
- Rieder, Bernhard (2004): Algorithmische Mediatoren: Software-Agenten als Gegenstand der Medienwissenschaften. In: *Medien-Journal*, 28. Jg., H. 1, S. 36–46.
- Rieder, Bernhard (2005): Networked Control: Search Engines and the Symmetry of Confidence. In: *International Review of Information Ethics*, 3. Jg., S. 26–32.
- Stöber, Rudolf (2008): Innovation und Evolution: Wie erklärt sich medialer und kommunikativer Wandel? In: Carsten Winter/Andreas Hepp/Friedrich Krotz (Hg.): *Theorien der Kommunikations- und Medienwissenschaft. Grundlegende Diskussionen, Forschungsfelder und Theorieentwicklungen*. Wiesbaden: VS Verlag für Sozialwissenschaften, S. 139–156.
- Stray, Jonathan (2011): A Computational Journalism Reading List. In: Jonathan stray.com. <http://jonathanstray.com/a-computational-journalism-reading-list>. Veröffentlicht am 31. Januar; zugegriffen am 16. Mai 2016.

Algorithmen, Big Data und ihre Rolle im Computational Journalism

- Van Dalen, Arjen (2012): The algorithms behind the headlines. How machine-written news redefines the core skills of human journalists. In: *Journalism Practice*, 6. Jg., H. 5–6, S. 648–658.
- Van der Kaa, Hille/Krahmer, Emil (2014): Journalist versus news consumer: The perceived credibility of machine written news. Research Paper, 2014 Computation + Journalism Symposium, Columbia University, New York.
- Vos, Tim P./Heinderyckx, François (2015): *Gatekeeping in Transition*. New York: Routledge.
- Weber, Wibke/Rall, Hannes (2013): We Are Journalists': Production Practices, Attitudes and a Case Study of the New York Times Newsroom. In: Wibke Weber/Michael Burmester/Ralph Tille (Hg.): *Interaktive Infografiken*. Wiesbaden: VS Verlag für Sozialwissenschaften, S. 161–172.
- Weeks, Lynn (2014): Media Law and Copyright Implications of Automated Journalism. In: *Journal of Intellectual Property and Entertainment Law*, 4. Jg., H. 1, S. 67–94.
- Weinacht, Stefan/Spiller, Ralf (2014): Datenjournalismus in Deutschland. Eine explorative Untersuchung zu Rollenbildern von Datenjournalisten. In: *Publizistik*, 59. Jg., H. 4, S. 411–433.
- Weischenberg, Siegfried/Hienzsch, Ulrich (1994): Die Entwicklung der Medientechnik. In: Klaus Merten/Siegfried J. Schmidt/Siegfried Weischenberg (Hg.): *Die Wirklichkeit der Medien. Eine Einführung in die Kommunikationswissenschaft*. Opladen: Westdeutscher Verlag, S. 455–480.
- Yarnall, Louise/Johnson, J.T./Rinne, Luke/Ranney, Michael Andrew (2008): How Post-Secondary Journalism Educators Teach Advanced CAR Data Analysis Skills in the Digital Age. In: *Journalism & Mass Communication Educator*, 63 Jg., H. 2, S. 146–164.
- Young, Mary L./Hermida, Alfred (2015): From Mr. and Mrs. Outlier To Central Tendencies. In: *Digital Journalism*, 3. Jg., H. 3, S. 381–397.
- Zion, Lawrie/Craig, David (2014): *Ethics for Digital Journalists. Emerging Best Practices*. New York: Routledge.

Zusammenfassung

With software automatically producing texts in natural language from structured data, the evolution of natural language generation (NLG) is changing traditional news production. The paper first addresses the question whether NLG is able to perform the functions of professional journalism on a technical level. A technological potential analysis therefore uncovers the technological limitations and possibilities of NLG, accompanied by an institutional classification following Weischenberg, Malik, and Scholl. Overall, NLG is explained within the framework of algorithmic selection and along its technological functionality. The second part of the paper focuses on the economic potential of NLG in journalism as well as indicating its institutionalization on an organizational level. Thirteen semi-structured interviews with representatives of the most relevant service providers detail the current market situation. Following Heuss, the development of the NLG market is classified into phases. In summary, although the market for NLG in journalism is still at an early stage of market expansion, with only a few providers and journalistic products available, NLG is able to perform tasks of professional journalism at a technical level. The analysis therefore sets the basis to analyze upcoming challenges for journalism research at the intersection of technology and big data.

Dörr, K. 2016. Mapping the Field of Algorithmic Journalism. *Digital Journalism*, 4(6), 700–722.

Copyright: <http://www.tandfonline.com>

The article is published as the version of record in *Digital Journalism*:
<http://www.tandfonline.com/doi/full/10.1080/21670811.2015.1096748>

MAPPING THE FIELD OF ALGORITHMIC JOURNALISM

Konstantin Nicholas Dörr

With software automatically producing texts in natural language from structured data, the evolution of natural language generation (NLG) is changing traditional news production. The paper first addresses the question whether NLG is able to perform the functions of professional journalism on a technical level. A technological potential analysis therefore uncovers the technological limitations and possibilities of NLG, accompanied by an institutional classification following Weischenberg, Malik, and Scholl. Overall, NLG is explained within the framework of algorithmic selection and along its technological functionality. The second part of the paper focuses on the economic potential of NLG in journalism as well as indicating its institutionalization on an organizational level. Thirteen semi-structured interviews with representatives of the most relevant service providers detail the current market situation. Following Heuss, the development of the NLG market is classified into phases. In summary, although the market for NLG in journalism is still at an early stage of market expansion, with only a few providers and journalistic products available, NLG is able to perform tasks of professional journalism at a technical level. The analysis therefore sets the basis to analyze upcoming challenges for journalism research at the intersection of technology and big data.

KEYWORDS: algorithmic journalism; automation; institution; market potential; natural language generation; news production; robot journalism; technology

Introduction: Automating News Production

Historically, computerization shows that software has been taking over routine tasks of journalists (Frey and Osborne 2013) and “[j]ournalism has always been shaped by technology” (Pavlik 2000, 229). With algorithms editing, aggregating, publishing, and distributing content, processes of media production and consumption are increasingly being automated (Broussard 2014; Diakopoulos 2014; Gillespie 2014; Mager 2012; Napoli 2014; Pavlik 2013).

As newspapers struggle for profitability, market share, journalistic reputation, and readers, recent developments in news production go along with progress in natural language generation (NLG), a subfield of natural language processing (for an overview, see Jones 2001). Terms like “robot journalism,” “automated journalism,” “algorithmic journalism,” or “machine-written journalism” dominate the media and scientific discourse (Anderson 2012, 2013). Due to the rising availability of digital and digitized data, NLG is defined as software and computer systems, which automatically produce human (natural) language from a computational representation of information (Reiter and Dale

2000). Although companies like Narrative Science or Automated Insights are already able to add and embed graphics or other media to their generated texts, these additions are not NLG in the understanding of computational linguistics as they rely on other technological premises (Reiter 2010). Therefore they are not part of this analysis. But of course, it is the interplay between NLG and these tools of visualization that enhances the variety and content diversity of journalistic products.

Presently, there are only a few studies that discuss the impact of NLG on journalism. For example, Petre (2013) and Coddington (2014) identify and analyze a “quantitative turn” in journalism in different concepts like data journalism, computational journalism, and computer-assisted reporting. On the other hand, Carlson (2014) discusses this development—using the term automated journalism—mainly in relation to labor and authority. Van Dalen (2012) conducted interviews with journalists to analyze their attitudes concerning the launch of a machine-written sports website. He demonstrates how this technology forces journalists to re-examine their skills. Latar (2015) notes the limits of NLG in a more universal understanding of the potential of algorithms, lacking an institutional discussion. With journalism offering “institutionalized solutions of communication problems” (Neuberger 1996, 12), the rising potential of NLG is also leading to rising user expectations (Bateman 2010, 640). Therefore, initial studies on the perceived quality and credibility of algorithmically generated texts show that differences between human and automated written texts are nearly undiscernible (Clerwall 2014; Van der Kaa and Krahmer 2014). Questions about “how decisions of inclusion and exclusion are made, what styles of reasoning are employed, whose values are embedded into the technology, and how they affect public understanding of complex issues” (Young and Hermida 2015, 384) also challenge NLG on an ethical level. Carlson (2014) therefore addresses these questions with the concept of “algorithmic authority” and Gillespie (2014) discusses this development using the term “algorithmic objectivity.” This overview shows that algorithms are entering the formerly sacred territory of “human” text production and call for research on various levels (Poynter 2014, 2015). Besides the institutional discussion of automated news by Napoli (2014) and Anderson (2012, 2013), communication science has not been able to supply a coherent model of NLG in journalism so far, as well as identifying the possibilities and limitations of this technology. Thus, this leads to the first of two central questions addressed in this article to examine the development at the “intersection of journalism and technology” (Lewis and Usher 2013, 603).

RQ1: Is NLG able to perform the functions of professional journalism on a technical level?

As a theoretical basis for an institutional classification of this technology in journalism, this paper answers the question concerning whether NLG is able to perform functions of professional journalism on a technical level. First, NLG and its technical functionality are explained within the framework of “Algorithmic Selection on the Internet” by an input-throughput-output model (I-T-O model; Latzer et al. 2014). Latzer et al. (2014) identify algorithmic selection as the technical-functional core of many successful software applications fulfilling social tasks (e.g. reducing transactions costs). Moreover, they outline a functional typology and subsume text creation applications as a part of it. As outlined above, this paper solely focuses on content creation applications on the basis

of NLG and its use in journalism, where they “touch deeply upon human areas of creativity and expression” (Latzer et al. 2014, 8) and point to a new professional role where journalists are “migrating from a direct to an indirect role” (Napoli 2014, 350).

To uncover the technical limitations and possibilities of NLG, a technological potential analysis is used as an analytical tool, following the potential analysis of online journalism by Neuberger (2001) and Wolf (2014) for applications in mobile journalism. Knowing that this analysis can be viewed from a technical and an institutional perspective (Wolf 2014), both positions are identified not as contrary but as two stages during the institutionalization of NLG in journalism. Based on Weischenberg, Malik, and Scholl (2006), it is assumed that journalism is constituted by three spheres: social, organizational, and professional. The analysis of the technological potential of NLG is discussed within these spheres, along with an analysis of the basic functions, codes, and norms of journalism and in the light of various practical and theoretical approaches to defining “journalism” (Meier 2011; Neuberger 2002; Neuberger and Kapern 2013).

Thus, this development is here conceptualized as Algorithmic Journalism and is defined as the (semi)-automated process of NLG by the selection of electronic data from private or public databases (input), the assignment of relevance of pre-selected or non-selected data characteristics, the processing and structuring of the relevant data-sets to a semantic structure (throughput), and the publishing of the final text on an online or offline platform with a certain reach (output). It is produced inside or outside an editorial office or environment along professional journalistic guidelines and values that meet the criteria of topicality, periodicity, publicity, and universality, and thus establishes a public sphere. The technology of NLG is furthermore identified as the central technical innovation that enables Algorithmic Journalism.

To outline the economic potential of NLG in journalism (for Big Data Journalism, see Lewis and Westlund 2014) and to discuss the process of institutionalization on the organizational sphere, the second part of the paper addresses and answers the question:

RQ2: Who are the most relevant service providers in NLG with journalistic orientation and how is the present state of the market?

After identifying the most relevant service providers in NLG, 13 semi-structured interviews form the basis of this analysis. The interviews were conducted between December 7, 2014 and September 2, 2015 via Skype (nine), telephone (one) or face to face (three), and lasted between 27 and 124 minutes. To gain insights in the development of the market and its journalistic orientation, this paper follows Heuss (1965) and classifies NLG into market phases. While publicly available information about the companies and their clients is limited, the interviews allow conclusions on the market’s potential and the use of the technology in professional journalism.

The methodological and conceptual approach of this paper is based on triangulation, following Flick (2008, 12) and Denzin (1970, 1989). By triangulation the combination of methods and theoretical perspectives can lead to an increase in knowledge about NLG in journalism and is suitable for the analysis of this new technological development as a “fully grounded interpretive research approach” (Denzin 1989, 246).

The technological potential analysis as well as the economic overview therefore set the basis to analyze upcoming challenges for journalism research at the intersection of technology and big data.

Algorithmic Selection and NLG

First, NLG is explained within the framework of algorithmic selection along an I-T-O model. Although NLG is not a new research field, having first appeared in the 1950s as a minor part of machine translation (McDonald 2010; Reiter 2010), it only became an independent research subfield in the 1980s. The processes of language generation steadily advanced due to growing data availability and the importance of statistical data analysis (Reiter 2010). A popular application of NLG was textual weather forecasting, e.g. FOG (Goldberg, Driedger, and Kittredge 1994) or SumTime (Reiter et al. 2005). Other NLG applications are based on sports (Robin and McKeown 1996), medical (Portet et al. 2007), financial (Kukich 1983) or engineering data (Yu et al. 2007). There are also narrative NLG applications to persuade or motivate (Reiter, Robertson, and Osman 2003) and entertain (Binstead and Ritchie 1997), leading to automated storytelling (Callaway and Lester 2002; Perez y Perez and Sharples 2004).

Media reports may discuss “robots” in journalism but more precisely, algorithms are used to generate text. Latzer et al. (2014, 4), based on Cormen et al. (2009), define algorithms “as a finite series of precisely described rules or processes to solve a problem” and as “a sequence of stages that transforms input through specified computational procedures (throughput) into output.” Algorithms are dynamic in nature as they are “constantly adjusted in efforts to improve their performance in accordance with specific criteria” (Napoli 2014, 344).

Depending on the purpose of the application Latzer et al. (2014) use this basic definition for the framework of “Algorithmic Selection on the Internet,” where a selection of elements from a basic set is processed according to certain rules, followed by a structuring and ranking of relevant information into an intended output. These two aspects—the selection and the assignment of relevance—are reflected in the functionality of the framework and its applications. This basic functionality can also be applied to applications of NLG as they operate similarly (Carstensen et al. 2010; McDonald 2010; Reiter 2010; Reiter and Dale 2000). By “articulating the specifications of a system through a rigorous examination drawing on domain knowledge, observation, and deduction to unearth a model of how that system works” (Diakopoulos 2014, 404), the mode of operation of NLG can be also illustrated by the design of an I-T-O model (Figure 1).

Besides this adaption of functionality, NLG is also able to fulfill promises of algorithmic selection. As will be shown in the technological potential analysis, NLG is able to reduce various kinds of transaction costs due to process automation (e.g., search and information costs) (Latzer et al. 2014, 29). Based on initial programming of the NLG software (e.g., predefined statistical rules), the software is therefore able to draw conclusions and identify interesting facts within the underlying set of data for text generation autonomously and much quicker than a “human” journalist can achieve.

The starting point for NLG and its journalistic use is a database, for example sports, financial, weather or traffic data (input), which is processed according to predefined linguistic and statistical rules (throughput) to a text (output) in natural language.

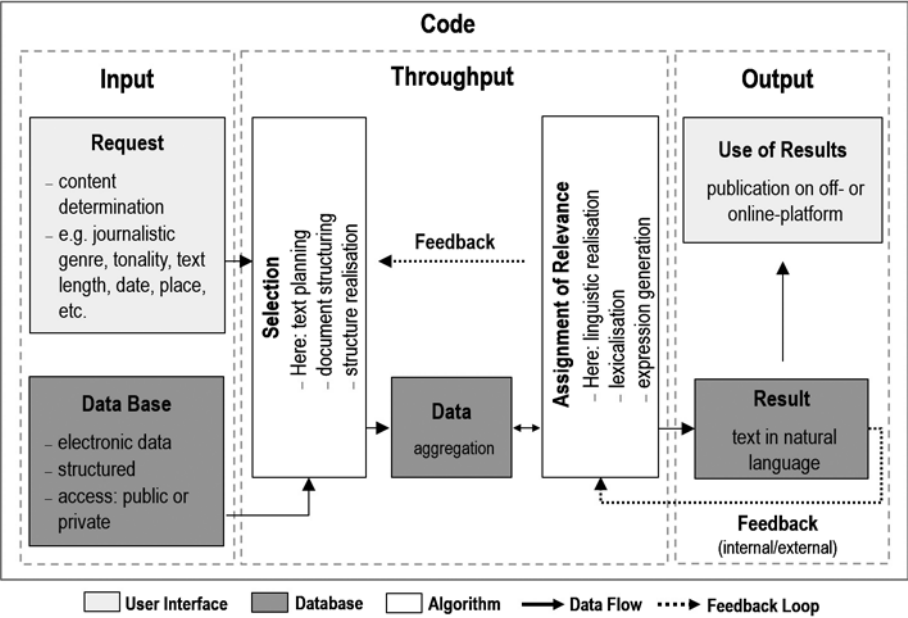


FIGURE 1
I-T-O model Algorithmic Journalism based on Latzer et al. (2014) and Reiter and Dale (2000)

A language is a set of strings over some alphabet, which is sometimes specified by grammars within a grammar framework (Pratt-Hartmann 2010, 55). Within this framework, any grammar recognizes a unique language, which has to be coded in order to generate texts in different languages. Thereby the process of generation is often divided into three stages: (1) document planning, (2) micro planning, and (3) realization. The following section refers to the research of Reiter and Dale (2000) and Reiter (2010) for a simplified understanding of the architecture of NLG for journalistic news production.

In general terms, the goal of *document planning* (input level) is to identify the information that is useful to the user or the intended output. The input to the document planner—structured data—is the input to the entire NLG system. This data, the “main ingredient” of NLG in journalism, can be accessed via public application programming interfaces (APIs) or via private databases (e.g., internal client data). For content-related text generation and therefore every journalistic product based on NLG, individual codes, rules, and dictionaries have to be coded and adapted. Thus, NLG systems only operate according to pre-set specific rules. This includes the linguistic creation process as well as the decision which hidden facts in the data should be processed and transformed into natural language. Thus, the final result is, apart from some linguistic adjustments, already almost definite. These pre-structured tasks are identified as a prerequisite at the input level/request level of Algorithmic Journalism (Reiter and Dale 2000, 49). This involves parameters such as text length, content/facts, journalistic form of presentation, theme, tonality, as well as the time and place of publication.

The core of this generation process is the throughput level (*micro planning*) that defines the input–output relationship (*text planning* and *realization*). Starting from the

input level (request), algorithms apply statistical operations to select elements from a basic data-set and assign relevance to them. The NLG system must decide which linguistic structures (words, syntax, sentences) should be used to communicate the desired information and in the *realization* (throughput) stage, it must decide which forms of words to use, and in which order they will appear. Feedback loops thereby indicate the human influence on the generation process which is optimized until the desired result is achieved. As Reiter (2010, 577) notes, this content generation process is complex and requires many decisions, including the *lexical choice* (choosing which content and words should be used to express domain concepts and data reference); *referring expressions* to identify domain entities; *syntactic choice* (choosing syntactic structures in generated sentences; and *aggregation* (choosing how many messages should be expressed in each sentence).

The result (output) is a text in natural language. After this generation process, the texts are published mainly automatically on online or offline news outlets.

Technological Potential Analysis of NLG

To answer the question whether NLG is able to perform the functions of professional journalism, a technological potential analysis is used as a suitable tool to discover the technical possibilities and limitations of a technology (Neuberger 2001, 92).

As pointed out, this analysis can be viewed from a technical and an institutional perspective (Wolf 2014), while both positions are identified not as contrary but as two stages during the institutionalization of NLG in journalism. Also Young and Hermida (2015, 383) point out that “the development and application of algorithms in journalism requires both a technological and sociological lens.” Therefore, NLG is regulated and influenced by the characteristics of the internet as an enabling-technology for data collection and text distribution via cloud-based solutions. From a sociological lens (see RQ2), for example, the integration of NLG products into organizational routines of professional media organizations have to be discussed (Kiefer 2010; Neverla 2001). As for now, this paper focuses on the technological lens first to analyze NLG from a primarily technology-centered view. Knowing that the analysis of its technical potential is only a necessary first step of a broader institutional analysis as expectations of media companies, the industry, and the public have to be taken into account (Wolf 2014, 70).

The technical perspective is illustrated according to the technological potential analysis by Neuberger (2001) for online journalism and Wolf (2014) for applications in mobile journalism. For this analysis, 10 technical potential dimensions are identified, derived from scientific literature focusing on NLG (Carstensen et al. 2010; McDonald 2010; Reiter 2010; Reiter and Dale 2000) and journalism (Neuberger 2001; Wolf 2014). As there is limited research on NLG and its use in journalism, additional technological characteristics of NLG were derived from the interviews.

The final potential dimensions therefore include established online characteristics like multimediality, additivity, topicality, selectivity, context sensitivity, and interactivity (Neuberger 2003, 57), as well as specific technical features of NLG like data availability, data quality, language diversity, and production routines (see Table 1).

TABLE 1
Analytical dimensions of the potential analysis of NLG derived from Neuberger (2001), Wolf (2014), Reiter and Dale (2000), McDonald (2010), Reiter (2010), Carstensen et al. (2010), and accompanying interviews with the service providers

Characterstics of the Internet	Characteristics of NLG
<ul style="list-style-type: none">- Multimediality- Interactivity- Additivity- Topicality- Selectivity- Context sensivity	<ul style="list-style-type: none">- Data availability- Data quality- Language diversity- Production routines

To frame this technological development for journalism research, this paper does not propose a new theory or definition of journalism but instead provides a comprehensive classification. Although there are many possible understandings of journalism as well as theoretical approaches (Löffelholz 2008; Quandt 2005, 23), it is agreed upon that journalism is a form of public communication (Quandt 2000, 484). In this paper, “journalism” is seen as “as a social system that enables society to observe itself, as it provides the public independently and periodically with information and issues that are considered newsworthy, relevant, and fact-based” (Weischenberg, Malik, and Scholl 2012, 207). Central to this theoretical understanding are media organizations as institutions, which fulfill specific functions for society (Weischenberg, Malik, and Scholl 2006, 347).

Analyzing NLG in journalism from an institutional perspective, this paper follows Weischenberg, Malik, and Scholl (2006) and argues that journalism is constituted through a social, organizational, and professional sphere. On a social level, journalism fulfills certain tasks by observing parts of society and provides the public with relevant information. This also includes fact-based products with a certain reach (Weischenberg, Malik, and Scholl 2006, 346) as well as products of special interest (Rühl 1980, 382). Journalism is traditionally produced on the organizational level within media organizations according to specific rules and routines or by other journalistic actors on a professional level (Wolf 2014, 66; Hohlfeld 2003, 127; Weischenberg, Malik, and Scholl 2006, 346). Furthermore, the principles of topicality, periodicity, publicity, and universality (Groth 1960, 360) are *not* linked to a technical artifact such as the newspaper and are also transferable to the general understanding of Algorithmic Journalism outlined above (Wolf 2014, 67). As long as the technology meets the prerequisites above, it is able to perform tasks and fulfill functions of professional journalism on a technical level (see Table 1).

This preliminary study of the technological potential dimensions serves as a base for the institutional classification afterwards.

Multimediality, Interactivity, and Additivity

While multimedia is the integration of different media for the communication and mediation of information, including text, photo, graphic, video, animations, and

audio (Meier 2002, 129), presently, NLG in the sense of computational linguistics is the generation of texts in human language (Reiter 2010). But as outlined above, it is the combination of different tools and other technologies that complement journalistic NLG products. While digital journalism in general is still primarily text-based and is only enriched with photographs, videos, and graphics (Quandt 2008, 140), NLG as the core technology constantly progresses and fulfills the prerequisite of multimodality. As there are already NLG systems that include hyperlinks (Reiter 2010, 595), therefore *interactivity* as the technological potential to enable consumer interaction is possible (Neuberger 2007, 43). The more customized the final text product, the more effective and usable it is for the reader or client. This is a major reason for the use of NLG systems in journalism (Bateman 2010, 635). Hypertextuality as a form of journalistic text design therefore enables “a non-linear, modular presentation of texts with multimedia elements” (Neuberger 2003, 63) and enables linking and distribution of already existing content (*additivity*). On the one hand, the text can be the final product, but may also be the starting point for further editing, adding additional information like quotes or images. The internet, therefore, acts as a carrier medium and distribution channel, while the consumption of this content is not linked to a specific device. Overall, NLG can be seen as a product extension of digital journalism.

Topicality

Publication speed is important in digital journalism and online content can be continuously updated if (live) data is available, thereby ignoring traditional production routines (Neuberger 2003, 60). Presently, structured data sources are hardly available for NLG (e.g., only for specific domains). The term topicality includes the speed of content production within media organizations as well as the relevance of a topic (Meier 2003, 253). The connectivity to databases (public or private, offline or online) therefore shortens the time for content production, as NLG operates independently after final programming. Depending on the availability, quality, and topicality of the data, as well as the technological infrastructure of the provider and the client, NLG is able to generate a large number of texts at any time. Thus, news is far more visible, for example, to search engine indexing. This is not to say that NLG systems do not produce errors as they are definitely able to do wrong reporting based on wrong coding or wrong data sources (Fox News 2015). This is also leading to ethical challenges. As, for example, the Associated Press (AP) has stopped monitoring every single generated text produced by Automated Insights for their earnings reports as it is too time consuming (Turi2 2015).

Selectivity and Context Sensitivity

Selectivity in the sense of social sciences focuses on the user and the potential of the system to personalize and select specific functions for the user (e.g., layout or content; Neuberger 2003, 61; 2007, 44). While NLG needs the technological prerequisites of the internet (e.g., cloud-based distribution of texts and data collection) for a broad integration into journalism, it is possible that users select special topics and journalistic niche products and personalize them based on their individual interests once texts are

generated. As Wolf (2014, 94) notes, content refers to the local, temporal, event-related, and interest-specific sphere of the user. This can be news on the favorite football club or the latest developments of the stock market. With NLG being *context-sensitive* and service providers heavily focusing on processing real-time data for text generation (e.g., stock market news), the audience is able to get the right information at the right time and place according to their individual interests (e.g., push messages in mobile use).

Data Availability, Data Quality, Language Diversity, and Production Routines

With data being the main resource, service providers have partnerships with media agencies and data brokers in terms of data acquisition. On the other hand, they process data already delivered and scraped by their clients. While the technological and linguistic capabilities of NLG are steadily increasing, the availability of data is the main issue. By now, companies and clients tend to scrape data via open APIs, however, they often do not disclose these origins. Highly structured and detailed data improve the quality of the generated texts. Today, NLG systems are even able to generate texts in multiple languages (Reiter and Dale 2000; also see Table 3). Moreover, the automation of NLG reduces transactions costs for search and information, and journalists are thus able to focus on other relevant tasks as the software is processing the data autonomously after coding.

Considering that NLG is a highly complex process in the field of computational linguistics, algorithms are not able to generate texts without human interference. But still, these new production routines are leading to shifts in journalistic roles. Although evolving journalistic work processes are not new, they often force “new tasks on reporters and editors alike” (Powers 2012, 27). I argue that the direct and active human element *during* the process of content creation is eliminated in Algorithmic Journalism. This is not to say that the human factor is eliminated from content creation altogether, because algorithms are themselves developed by humans. The point is that the individual journalist in NLG is changing to a more indirect role (Napoli 2014) *before, during, and after* text production; as, for example, source selection (input), fact checking, the actual writing (throughput—both depending on coding) and distribution (output) are automated and pre-selected in Algorithmic Journalism. Journalistic work and knowledge now has to comprise skills like programming even more.

The effectiveness of modern NLG systems is predicated on domain specialization, i.e., they “restrict themselves to a specialized area of discourse with a very focused audience and stipulated content, thereby reducing the options for word choice and syntactic style to a manageable set” (McDonald 2010, 126). Therefore, the development of the semantic web (Berners-Lee, Hendler, and Lassila 2001), which focuses on the large availability of structured knowledge domains, offers vast opportunities for the further development of NLG systems. To date, NLG and its use in journalism is feasible in clearly defined domains such as weather, traffic, finance, or sports with good input data availability. However, once the domain and underlying set of world knowledge expands, NLG struggles with ambiguity of words and phrases (e.g., ball: leather ball or dance?). This is why companies focus on journalistic products of the sports and

TABLE 2
 Technical premises, possibilities, and limitations of NLG in journalism

Premises	Possibilities	Limitations
<ul style="list-style-type: none"> - Data availability - Data quality 	<ul style="list-style-type: none"> - Production routines; e.g., reduce transactions costs for search and information - Additivity - Selectivity; e.g., conquer niche markets and special-interest content - Context sensitivity - Language diversity - Topicality; e.g., information and distribution speed; visibility of content (SEO) - Multimediality 	<ul style="list-style-type: none"> - Production routines; e.g., world knowledge (special domains and contexts) - Production routines; e.g., reflection of content and assessment of facts, reasoning - Production routines; e.g., narration (arc of suspense) - Production routines; e.g. source checking - Interactivity

financial domain where processable and structured data (e.g., .xls, .csv, .xml, .json) are available (see Table 3). Data for text generation has to be available in a scalable amount as programming and individualization of journalistic products are cost-intensive (e.g., pilot projects of NLG in journalism can cost up to 25,000 Euro depending on the programming effort; interview, Claude de Loupy).

Another limitation is that NLG is not able to draw conclusions from contradictory data. For example, there is no source checking and the databases have to be very reliable and well-maintained (Carstensen et al. 2010).

Independent journalistic interpretation and reflection, for example, of political issues, poses major challenges for NLG. Journalistic narration (e.g., arc of suspense) is also severely limited (Reiter 2010). NLG in journalism is thus more of a starting point and information can be added to the generated text (e.g., quotes from stakeholders, experts, etc.).

Table 2 sums up the extent of the premises, possibilities, and limitations of NLG derived from the potential analysis.

Building on the technical perspective, the following section discusses the first research question concerning NLG’s ability to fulfill the functions of professional journalism on a technical level. This is based on the institutional understanding of Weischenberg, Malik, and Scholl (2006).

Relevant Information, Including Special-interest Topics

With NLG already able to produce articles for different sports (e.g., ice hockey, basketball, American football, football), weather, traffic, finance, and celebrity news, it also produces relevant information for special-interest topics (e.g., NCAA college sports such as baseball and lower-division basketball and football; see the results of the interviews in Table 3). This adds value to traditional news coverage by offering coverage that was previously not profitable.

Reach

The early integration of NLG into the daily operation of media organizations is already able to generate a certain reach (see Table 3). For example, the major wire service AP uses this technology to generate earning reports for their clients in the AP Style Guide. Such reports are labeled “This story was generated by Automated Insights using data from Zacks Investment Research.” AP also invested in Automated Insights, a NLG service provider, via the private equity firm Vista Equity Partners (AP 2015).

Information Produced Within Media Organizations/Newsrooms

News is already produced within media organizations and the newsroom (e.g., the fine dust monitor by *Berliner Morgenpost* was developed by a team of five journalists and programmers). NLG services and complete articles can also be bought externally from service providers. As long as these services are labeled and communicated within the brand of the media organization, this fulfills the requirements of a production within organization.

Professional Journalistic Actors

To date, media organizations are already testing and using NLG services in journalism (see Table 3), evaluating whether products match the standards and quality of their brand and how journalists and users react. These case studies also guarantee the influence of the professional journalistic actor on the individual product, which is tailored to the needs of the client and the audience. Therefore, the software training guarantees the matching of the individual writing style and the expected output. No media organization integrates automated news without serious testing, as media organizations have more to lose than just readers (e.g., credibility of their brand). At this stage of product development, the journalistic actor heavily influences the processing and final version of the text. But questions about “new” actors arise if service providers or other players outside professional journalism (e.g., data brokers) decide to reach an audience without professional journalistic intermediaries.

Information Produced According to Specific Rules and Routines

In general, “the use of algorithms for media production contains an editorial logic based on the socially situated choices of media professionals” (Young and Hermida 2015, 384). The influence of the journalist in NLG is visible as the texts are generated according to different rules and routines specific to the media organization. For example, the adjustment of the AP earning reports matching the AP Style Guide took nearly one year before launch (American Society of Business Press Editors 2015). This also fits the requirements of an institutionalized journalism. Media organizations have to be transparent about the data sources and the labeling of the texts. This is closely linked to algorithmic transparency, which Stavelin (2013) argues is borrowed from professional journalistic values. The legal challenges of copyright and accountability are quite

unclear and the role of the journalist is in flux. Consequently, the routines of news production are changing.

Principle of Topicality, Periodicity, Publicity, and Universality

Argued from a technological-institutional perspective, NLG is able to perform the principle of topicality (e.g., Quakebot reports immediately after the earthquake; sports results after or during the games). The principle of periodicity is fulfilled as texts can be generated regularly—if data is available. When news is integrated within a special journalistic offering, automated texts are accessible publicly via the internet. With NLG also covering a number of different topics (sports, entertainment, finance, weather), it also fulfills the principle of universality, to date with a limitation to mostly special-interest content.

Observation of Society

The above conditions enable the observation of society within different topics, although there are technical limitations (reasoning, reflection, and interpretation). Despite this, NLG is able to add relevant information within special topics to the traditional news coverage in order to fulfill journalistic tasks of orientation and opinion formation. Anderson (2011, 541) suggests that Algorithmic Journalism “might represent the most recent, and thus most unsettling, model for both communication and democracy” and Coddington (2014, 332) notes that forms of quantitative journalism like Algorithmic Journalism have “great potential to broaden journalism’s ability to make democratic institutions more responsive and legible to the public.”

Thus, the analysis of the technological potential as well as the institutional discussion show that overall, NLG is able to fulfill certain tasks of professional journalism on a technical level. Therefore, NLG is conceptualized on a technical level as Algorithmic Journalism and is defined as shown above.

The Emerging Market for Algorithmic Journalism

After the technological potential analysis, which outlined that NLG is able to perform tasks of professional journalism, a review of available market data gives insights about the present state of the market for Algorithmic Journalism indicating the process of the institutionalization of NLG on an organizational level. The classification into market phases therefore allows conclusions on its potential growth as there is a connection between market phases and market structures. Heuss (1965) notes that markets go through five ideal-type stages during their development: the experimental phase where the product is invented, developed, and launched; the expansion phase characterized by exponential growth; the maturity phase where growth diminishes; the stagnation phase with nearly no growth; and finally the regression phase where the market reverts back. These phases are often characterized by high concentration and by temporary monopolies of innovators and early movers (Latzner et al. 2014, 14).

While NLG providers were not able to enter the market long term until the end of 2010 (Reiter 2010), there are companies and start-ups that enter the market with different products based on NLG. A literature analysis first identified 13 leading service providers with NLG technology as their core business. Given the very limited information available about such companies—particularly on the development and offer of *journalistic* NLG products—interviews were conducted to complement the market overview. These semi-structured expert interviews, which lasted between 27 and 124 minutes, were conducted with the following companies (for a complete list of all interviewees, see Appendix A): Narrative Science (United States)—via Skype; Automated Insights (United States)—via Skype; OnlyBoth (United States)—via Skype; Linguastat (United States)—via Skype; Retresco (Germany)—face-to-face; Aexea (Germany)—face-to-face; Text-on (Germany)—face-to-face; Textomatic (Germany)—via telephone; 2txt (Germany)—via Skype; Syllabs (France)—via Skype; Labsense (France)—via Skype; YSEOP (France/United States/United Kingdom)—via Skype; and Arria (United Kingdom)—via Skype.

As a first result, 10 out of 13 providers offer products with possible applications in journalism (see Table 3). In order to draw a holistic picture of the market, all service providers were asked about the legal form of the company, the founding year, external funding, employees, languages of text generation, topics covered for journalistic use, products available, journalistic clients, as well as potential competitors. First, no additional competitors were named indicating that the market of NLG and especially for its application in journalism is served by only a few (see Table 3).

With Tencent, a Chinese social and gaming service provider, there is already a new competitor rising. Unfortunately, Tencent could not be part of the interviews due to the submission deadline of this paper but was added to the analysis in order to be as up-to-date as possible. Via Dreamwriter, a Tencent-designed NLG tool, they published a business report written in Chinese covering basic financial news (*South Morning China Post* 2015).

As the companies differ in size, assets, and product portfolio, they decrease their dependency on one market segment as they also offer NLG solutions for, for example, e-commerce, finance, oil industry, healthcare, or the farming industry. This is likely due to the complexity of NLG, the limited availability of data, the time-consuming individualization of journalistic products due to high-quality standards in journalism, and the general view that journalistic products alone are hardly profitable. This complexity also forces media organizations to buy most of these services externally as they lack newsroom infrastructures and financial resources for development.

Considering the company's size, clients, and products already launched, Automated Insights and Narrative Science are the driving forces of Algorithmic Journalism in the United States. In Germany, these are Retresco and Aexea, in France, Syllabs, and in the United Kingdom, Arria. Other service providers like YSEOP (France/United States/United Kingdom), Linguastat (United States), and OnlyBoth (United States), who were also interviewed, do not offer products and NLG services for journalistic use and are therefore excluded from further analysis. While the majority of the providers do not receive external funding, Automated Insights, Narrative Science, Arria, and Labsense have external investors. The other companies are privately held. As the market is in

TABLE 3
Market of service providers of NLG with focus on journalism *

Company	Country	Legal form	Founding year	Launch, NLG software	External funding	Employees	Languages (to date)	Topics covered for journalistic use	Journalistic products	Journalistic clients
Automated Insights	US	Inc., subsidiary of Stats LLC	2007	2007, branded as Wordsmith in 2014	About \$10.8 million from 11 investors before being acquired by Vista Equity Partners	About 40	(1) ENG	Finance Sports	Corporate earnings stories (AP); NCAA college sports (AP) (in development); Yahoo Sports Fantasy Football	AP (US); Yahoo! (US)
Narrative Science	US	Inc.	2010	2011, Quill	About \$32.4 million from seven investors	About 80	(1) ENG	Finance Sports	Earnings estimates of stock market companies (Forbes); Sports statistics (e.g., basketball, American football, softball, baseball) (launched)	Forbes (US); Big Ten Network (US); Game Changer (US); 5–10 signed contracts with US media (not public)
Aexea	GER	GmbH	2001	2009, AX Semantics	No	About 42	(12) ENG, GER, FR, ESP, NL, DNK, SWE, NOR, IT, IDN, PRT, CHN	Sports Entertainment Finance Weather	Match announcements for all European soccer leagues (in German); American Football; Stock exchange reports; Celebrity football news (in German)	Five media clients (non-disclosure agreement) + Sports-Information-Service (SID) (GER)

(Continued)

TABLE 3. (Continued)

Company	Country	Legal form	Founding year	Launch, NLG software	External funding	Employees	Languages (to date)	Topics covered for journalistic use	Journalistic products	Journalistic clients
Text-on	GER	GmbH	2013	2014, Text-on	No	About six, not payed	(1) GER	Finance	Pilot project in the financial sector (in development) Berliner Morgenpost; Share price descriptions	Berliner Morgenpost (GER) Finanzen100.de (GER); Finanzen100.de (GER)
2txt NLG	GER	UG	2013	2013, 2txt	No	About five	(1) GER	Finance Sports	Football product (in development); Share price descriptions (in development)	Beginning of negotiations
Retresco	GER	GmbH	2008	2013, Rtr text engine	No	About 27	(1) GER	Sports	Preliminary reports of football games in lower German leagues (Kreisklasse)	Neue Osnabrücker Zeitung; Weserkurier; Radio Hamburg FussFreunde; Rheintussball; Goekick.info; Fubanews.org
Textomatic	GER	AG	2015	2015, Textomatic	No	About five	(6) GER, ENG, ESP, NL, FR, ITA	Sport Finance Traveling Weather	Football; Stock exchange reports; Travel advices; Personalized weather reports	two media clients (Handelsblatt and one regional newspaper)
Syllabs	FR	LLC	2006	2012, Data2content	No	About 11	(3) FR, ESP, ENG	Politics	Project on departmental elections 2015 in France	Le Monde (FR)

(Continued)

TABLE 3. (Continued)

Company	Country	Legal form	Founding year	Launch, NLG software	External funding	Employees	Languages (to date)	Topics covered for journalistic use	Journalistic products	Journalistic clients
Labsense	FR	SAS	2011	2013, ScribL	Less than \$565,000	About six	(3) FR, ENG, GER	Economy	Project on local news in France (in development)	Beginning of negotiations
Arria	UK	PLC	2011	2012, Arria NLG Engine	About \$36 million from shares	About 50	(1) ENG	Weather	Weather report module for two regions in Europe (UK and Germany)	MeteoGroup (UK)
Tencent	CHN	Ltd.	1998	2015, Dreamwriter	No information	No information	(1) CHN	Finance	CPI report on China's growth	No information

*COUNTRY ABBREVIATIONS: US, UNITED STATES; GER, GERMANY; UK, UNITED KINGDOM; CHN, CHINA. LANGUAGE ABBREVIATIONS: ENG, ENGLISH; GER, GERMAN; FR, FRENCH; ESP, SPANISH; NL, DUTCH; DNK, DANISH; SWE, SWEDISH; NOR, NORWEGIAN; IT, ITALIAN; IDN, INDONESIAN; PRT, PORTUGUESE; CHN, CHINESE.

constant flow, even companies with long-term experience in the field of NLG, like Arria, are quickly struggling for survival if clients terminate their contracts (e.g., Arria and Shell Exploration & Production Company; Arria NLG 2015). The investments by external business angels are not an indicator of market success. It is more of a vision of the technology's potential. The interviewees refused to give details regarding the revenue of the companies. Although service providers report that they are constantly negotiating with well-known media organizations, only a few products have been launched officially—those that are available mainly serve financial and sports reporting.

The primary reason why these companies focus on journalistic products for the sports and financial domain is that processable and structured data are available in a scalable amount. Additionally, the language, rules, and settings can be easily defined in terms of programming in a discrete framework and domain.

As a result of the interviews, Table 3 shows the market and gives an overview of the potential and core uses of Algorithmic Journalism. With few service providers, limited and resembling journalistic products available, Algorithmic Journalism is likely found either in an experimental market phase or in an early stage of the market expansion phase. But the integration of these products within professional media organizations already indicates a starting institutionalization on an organizational sphere (Weischenberg, Malik, and Scholl 2006).

Conclusion

Based on the framework of algorithmic selection and technological potential analysis, this paper identified the technological possibilities and limitations of NLG in journalism. Its journalistic application was conceptualized as Algorithmic Journalism with NLG serving as the central technical innovation enabling it. The described technical limitations as well as the dependence of NLG on structured data does not change the fact that NLG is able to perform institutionalized tasks of professional journalism on a technological level.

This mainly technological discussion therefore sets the basis to analyze upcoming challenges for journalism research at the intersection of big data like the shift of journalistic roles before, during, and after content production, ethical and legal challenges in news production, or questions about new intermediaries in journalism as the dependency on data grows.

Because news production remains a for-profit business, it is evident that publishers, on the one hand, seek to reduce costs (Kiefer 2001, 33; Stavelin 2013, 27). On the other hand, they are also desperately looking for new journalistic products and ways to satisfy their audience. The interviews and the classification into market phases show that the market for Algorithmic Journalism is still relatively small. But service providers already pressure traditional media organizations (AI 2015a, 2015b). The integration of Algorithmic Journalism into the portfolio of media organizations also indicates a starting institutionalization of Algorithmic Journalism on an organizational level. As the costs of NLG systems are low compared to human journalists, Algorithmic Journalism can be profitable for special-interest domains in the long tail, and also due to the possibility of generating news in multiple languages reach a broader audience and new markets.

ACKNOWLEDGEMENTS

The author would like to thank Michael Latzer and Natascha Just as well as the two anonymous reviewers for their helpful and constructive comments.

DISCLOSURE STATEMENT

No potential conflict of interest was reported by the author.

FUNDING

The research was supported by a research grant of the Swiss National Science Foundation (SNF).

REFERENCES

- Al. 2015a. "Robot Reporters: How the AP Embraced Data Automation | Strata + Hadoop World." <https://www.youtube.com/watch?t=31&v=IBzXiMiwQQs>.
- Al. 2015b. "When Robots Write the News, What Will Humans Do? | SXSW Interactive." <https://www.youtube.com/watch?v=0LqLHAtLE-c>.
- Anderson, Christopher W. 2011. "Understanding the Role Played by Algorithms and Computational Practices in the Collection, Evaluation, Presentation, and Dissemination of Journalistic Evidence." Paper presented at the 1st Berlin Symposium on the Internet and Society, Berlin, Germany, October 25–28.
- Anderson, Christopher W. 2012. "Towards a Sociology of Computational and Algorithmic Journalism." *New Media and Society* (15) 7: 1005–1021.
- Anderson, Christopher W. 2013. "What Aggregators Do: Towards a Networked Concept of Journalistic Expertise in the Digital Age." *Journalism* (14) 8: 1008–1023.
- AP. 2015. "Private Equity Firm Vista Buying Automated Insights." <http://www.ap.org/Content/AP-In-The-News/2015/Private-equity-firm-Vista-buying-Automated-Insights>.
- Arria, N. L. G. 2015. "Termination of Contract and Trading Update." *Email from Stuart Rogers, CEO Arria NLG* 30 (4): 2015.
- American Society of Business Press Editors. 2015. <http://www.asbpe.org/blog/2015/05/03/aps-tom-kent-the-time-has-arrived-for-robotics-journalism-ethical-checklist/>.
- Bateman, John. 2010. "Angewandte natürlichsprachliche Generierungs- und Auskunftssysteme [Applied Systems for Generation and Information Based on Natural Language]." In *Computerlinguistik und Sprachtechnologie. Eine Einführung* [Computational Linguistics and Speech Technology], edited by Kai-Uwe Carstensen, Christian Ebert, Cornelia Ebert, Susanne Jekat, Ralf Klabunde and Hagen Langer, 633–641. Heidelberg: Spektrum.
- Berners-Lee, Tim, James Hendler, and Ora Lassila. 2001. "The Semantic Web". *Scientific American* (284) 5: 24–30.
- Binstead, Kim, and Graeme Ritchie. 1997. "Computational Rules for Punning Riddles." *Humor. International Journal of Humor Research* 1 (10): 25–76.

- Broussard, Meredith. 2014. "Artificial Intelligence for Investigative Reporting. Using an Expert System to Enhance Journalists' Ability to Discover Original Public Affairs Stories." *Digital Journalism* (online first), <http://www.tandfonline.com/doi/abs/10.1080/21670811.2014.985497>.
- Callaway, Charles, and James Lester. 2002. "Narrative Prose Generation." *Artificial Intelligence* 2 (139): 213–252.
- Carlson, Matt. 2014. "The Robotic Reporter. Automated Journalism and the Redefinition of Labor, Compositional Forms, and Journalistic Authority." *Digital Journalism* 3 (3): 416–431.
- Carstensen, Kai-Uwe, Christian Ebert, Cornelia Ebert, Susanne Jekat, Ralf Klabunde, and Hagen Langer. 2010. *Computerlinguistik und Sprachtechnologie. Eine Einführung* [Computational Linguistics and Speech Technology. An Introduction]. Heidelberg: Spektrum.
- Clerwall, Christer. 2014. "Enter the Robot Journalist. Users' Perceptions of Automated Content." *Journalism Practice*, (8) 5: 519–531.
- Coddington, Mark. 2014. "Clarifying Journalism's Quantitative Turn. A Typology for Evaluating Data Journalism, Computational Journalism, and Computer-assisted Reporting." *Digital Journalism* 3 (3): 331–348.
- Cormen, Thomas H., Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. 2009. *Introduction to Algorithms*. Cambridge, Mass: MIT Press.
- Denzin, Norman K. 1970. *The Research Act*. Chicago, IL: Aldine.
- Denzin, Norman K. 1989. *The Research Act*. Englewood Cliffs, NJ: Prentice Hall.
- Diakopoulos, Nicholas. 2014. "Algorithmic Accountability. Journalistic Investigation of Computational Power Structures." *Digital Journalism* 3 (3): 398–415.
- Flick, Uwe. 2008. *Triangulation. Eine Einführung* [Triangulation. An Introduction]. Wiesbaden: VS Verlag.
- Fox News. 2015. "Correction: Earns-Graham Holdings Story". <http://www.foxnews.com/us/2015/08/07/correction-earns-graham-holdings-story/>.
- Frey, Benedikt Carl, and Michael A. Osborne. 2013. "The Future of Employment: How Susceptible Are Jobs to Computerisation?" Working Paper, University of Oxford, http://www.oxfordmartin.ox.ac.uk/downloads/academic/The_Future_of_Employment.pdf.
- Gillespie, Tarleton. 2014. "The Relevance of Algorithms." In *Media Technologies. Paths Forward in Social Research*, edited by Tarleton Gillespie, Pablo Boczkowski, and Kirsten Foot, 167–194. London: MIT Press.
- Goldberg, Eli, Norbert Driedger, and Richard Kittredge. 1994. "Using Natural-language Processing to Produce Weather Forecasts." *IEEE Expert* 2 (9): 45–53.
- Groth, Otto. 1960. *Die unerkannte Kulturmacht. Grundlagen der Zeitungswissenschaft (Periodik)* [The Unrecognized Cultural Power. Basics of Newspaper Research]. Berlin: Walter de Gruyter.
- Heuss, Ernst. 1965. *Allgemeine Markttheorie* [General Market Theory]. Tübingen: Mohr.
- Hohlfeld, Ralf. 2003. *Journalismus und Medienforschung. Theorie, Empirie, Transfer* [Journalism and Media Research. Theory, Empirical Research, Transfer]. Konstanz: UTB.
- Jones, Karen Sparck. 2001. "Natural Language Processing. A Historical Review." Working Paper, University of Cambridge. <http://www.cl.cam.ac.uk/archive/ksj21/histdw4.pdf>.
- Kiefer, Marie-Luise. 2001. *Medienökonomik* [Media Economics]. München: Oldenbourg.
- Kiefer, Marie-Luise. 2010. *Journalismus und Medien als Institutionen* [Journalism and Media as Institutions]. Konstanz: UVK.

- Kukich, Karen. 1983. "Design Andimplementation of a Knowledge-Basedreport Generator." Proceedings of 21st Annual Meeting of the Association for Computational Linguistics, 145–150.
- Latar, Noam Lemelshtrich. 2015. "The Robot Journalist in the Age of SocialPhysics: The End of Human Journalism?" In *The Economics of Information, Communication, and Entertainment. the Impacts of Digital Technology in the 21st Century*, edited by Gali Einav, 65–80. Wiesbaden: Springer.
- Latzer, Michael Hollnbuchner, Katharina Just, Natascha Saurwein, Florian. 2014. *The Economics of Algorithmic Selection on the Internet*. Zurich: University of Zurich. http://www.mediachange.ch/media//pdf/publications/Economics_of_algorithmic_selection_WP.pdf.
- Lewis, Seth C., and Oscar Westlund. 2014. "Big Data and Journalism. Epistemology, Expertise, Economics, and Ethics." *Digital Journalism* 3 (3): 447–466.
- Lewis, Seth C., and Nikki Usher. 2013. "Open Source and Journalism: Toward New Frameworks for Imagining News Innovation." *Media, Culture and Society* 35 (5): 602–619.
- Löffelholz, Martin. 2008. "Heterogeneous—Multidimensional—Competing: Theoretical Approaches to Journalism—An Overview." In *Global Journalism Research*, edited by Martin Löffelholz and David Weaver, 15–27. Hoboken, NJ: John Wiley & Sons.
- Mager, Astrid. 2012. "Algorithmic Ideology." *Information, Communication and Society* (15) 5: 269–787.
- McDonald, David D. 2010. "Natural Language Generation." In *Handbook of Natural Language Processing*, edited by Nitin Indurkha and Fred J. Damerau, 121–141. Chapman and Hall: Boca Raton, FL.
- Meier, Klaus. 2002. "Neue Journalistische Formen [New Journalistic Genres]." In *Internet-Journalismus* [Internet Journalism], edited by Klaus Meier, 21–172. Konstanz: UVK.
- Meier, Klaus. 2003. "Qualität Im Online-Journalismus [Quality in Online-Journalism]." In *Qualität Im Journalismus. Grundlagen, Dimensionen, Praxismodelle* [Quality in Journalism. Basics, Dimensions, Practical Models], edited by Hans-Jürgen Bucher and Klaus-Dieter Altmeyden, 247–266. Wiesbaden: Westdeutscher Verlag.
- Meier, Klaus. 2011. *Journalistik* [Journalism]. Konstanz: UTB.
- Napoli, Philip M. 2014. "Automated Media: An Institutional Theory Perspective on Algorithmic Media Production and Consumption." *Communication Theory* (24) 3: 340–360.
- Neuberger, Christoph. 1996. *Journalismus als Problembearbeitung Objektivität und Relevanz in der öffentlichen Kommunikation* [Journalism as Problem Processing. Objectivity and Relevance in Public Communication]. Konstanz: UVK.
- Neuberger, Christoph. 2001. *Journalismus im Internet. Theoriekontext und empirische Exploration* [Journalism on the Internet. Theoretical Context and Exploration]. *unveröffentlichte Habilitationsschrift*. Eichstätt: Universität Eichstätt-Ingolstadt.
- Neuberger, Christoph. 2002. "Vom Unsichtbarwerden Des Journalismus Im Internet [Journalism: Becoming Invisible on the Internet]." In *Innovationen Im Journalismus* [Innovations in Journalism]. *Forschung für die Praxis*, edited by Ralf Hohfeld, Klaus Meier, and Christoph Neuberger, 25–68. Münster: LIT Verlag.
- Neuberger, Christoph. 2003. "Zeitung Und Internet: Über Das Verhältnis Zwischen Einem Alten Und Einem Neuen Medium [Newspaper and Internet: About the Relationship of an old and new Medium]." In *Online—Die Zukunft Der Zeitung?* [Online - The Future of the Newspaper?] edited by Christoph Neuberger and Jan Tonnemacher, 16–110. Wiesbaden: Westdeutscher Verlag.

- Neuberger, Christoph. 2007. "Interaktivität, Interaktion, Internet [Interactivity, Interaction, Internet]." *Publizistik* 52 (1): 33–50.
- Neuberger, Christoph, and Peter Kapern. 2013. *Grundlagen Des Journalismus* [Basics of Journalism]. Wiesbaden: VS Verlag.
- Neverla, Irene. 2001. "Das Netz—Eine Herausforderung Für Die Kommunikationswissenschaft [The Internet: Challenges for Communication Science]." In *Kommunikationskulturen zwischen Kontinuität Und Wandel. Universelle Netzwerke Für Die Zivilgesellschaft* [Cultures of Communication between Continuity and Change], edited by Ursula Maier-Rabler und Ulrich Latzer, 29–46. Konstanz: UVK.
- Pavlik, John. 2000. "The Impact of Technology on Journalism." *Journalism Studies* (1) 2: 229–237.
- Pavlik, John. 2013. "Innovation and the Future of Journalism." *Digital Journalism* 1 (2): 181–193.
- Perez y Perez, Rafael, and Mike Sharples. 2004. "Three Computer-Based Models of Storytelling: BRUTUS, MINSTREL, and MEXICA." *Knowledge-Based Systems* 17 (1): 15–29.
- Petre, Caitlin. 2013. "A Quantitative Turn in Journalism?" *Tow Center for Digital Journalism*, October 30. <http://towcenter.org/blog/a-quantitative-turn-in-journalism/>.
- Portet, François, Ehud Reiter, Jim Hunter, and Somayajulu Sripada. 2007. "Automatic Generation of Textual Summaries from Neonatal Intensive Care Data." Proceedings of the 11th Conference on Artificial Intelligence in Medicine (AIME 2007), 227–236.
- Powers, Matthew. 2012. "In Forms That Are Familiar and Yet-to-be Invented: American Journalism and the Discourse of Technologically Specific Work." *Journal of Communication Inquiry* 36 (1): 24–43.
- Poynter. 2014. "L.a. times Reporter Talks about His Story-writing 'Quakebot'." <http://www.poynter.org/news/mediawire/243744/l-a-times-reporter-talks-about-his-story-writing-quakebot/>.
- Poynter. 2015. "AP Will Use Software to Write NCAA Game Stories." <http://www.poynter.org/news/mediawire/324601/ap-will-use-software-to-write-ncaa-game-stories/>.
- Pratt-hartmann, Ian. 2010. "Computational Complexity in Natural Language." In *Handbook of Computational Linguistics and Natural Language Processing*, edited by Alexander Clark, Chris Fox and Shalom Lappin, 43–73. Oxford: Wiley-Blackwell.
- Quandt, Thorsten. 2000. "Das Ende des Journalismus? Online-Kommunikation als Herausforderung für die Journalismusforschung [The End of Journalism? Online Communication as a Challenge for Journalism Research]." In *Theorien Des Journalismus. Ein Diskursives Handbuch* [Theories of Journalism], edited by Martin Löffelholz, 483–509. Wiesbaden: Westdeutscher Verlag.
- Quandt, Thorsten. 2005. *Journalisten im Netz Eine Untersuchung journalistischenhandelns in Online-Redaktionen* [Journalists on the Internet. A Study about Journalistic Acting in Online-Newsrooms]. Wiesbaden: VS Verlag.
- Quandt, Thorsten. 2008. "Neues Medium, Alter Journalismus? Eine vergleichendeinhaltsanalyse tagesaktueller Print- Und Online-Nachrichtenangebote [New Medium, Old Journalism? A Comparative Content Analysis of Print and Online News Services]." In *Journalismus Online—Partizipation Oder Profession?* [Journalism Online - Participation or Profession?] edited by Thorsten Quandt and Wolfgang Schweiger, 131–155. Wiesbaden: VS Verlag.

- Reiter, Ehud. 2010. "Natural Language Generation." In *The Handbook of Computational Linguistics and Natural Language Processing*, edited by Alexander Clark, Chris Fox and Shalom Lappin, 574–598. Oxford: Wiley-Blackwell.
- Reiter, Ehud, and Robert Dale. 2000. *Building Natural Language Generation Systems*. Cambridge: Cambridge University Press.
- Reiter, Ehud, Roma Robertson, and Liesl Osman. 2003. "Lessons from a Failure: Generating Tailored Smoking Cessationletters." *Artificial Intelligence* 144 (1–2): 41–58.
- Reiter, Ehud, Somayajulu Sripada, Jim Hunter, and Jin Yu. 2005. "Choosing Words in Computer-Generated Weather Forecasts." *Artificial Intelligence* 167 (1–2): 137–169.
- Robin, Jacques, and Kathleen R. McKeown. 1996. "Empirically Designing Andevaluating a New Revision-Based Modelfor Summary Generation." *Artificial Intelligence* 85 (1–2): 135–179.
- Rühl, Manfred. 1980. *Journalismus und Gesellschaft* [Journalism and Society]. *Bestandsaufnahme und Theorieentwurf*. Mainz: v. Hase & Koehler.
- South China Morning Post. 2015. "End of the Road for Journalists? Tencent's Robot Reporter 'Dreamwriter' Churns out Perfect 1,000-Word News Story - in 60 Seconds." <http://www.scmp.com/tech/china-tech/article/1857196/end-road-journalists-robot-reporter-dreamwriter-chinas-tencent>.
- Stavelin, Eirik. 2013. *Computational Journalism: When Journalism Meets Programming*. Norway: University of Bergen.
- Turi2. 2015. "AP Setzt Auf Roboterjournalismus [AP focuses on Robot Journalism]." <http://www.turi2.de/aktuell/ap-setzt-auf-roboterjournalismus/>.
- Van Dalen, Arjen. 2012. "The Algorithms behind the Headlines. How Machine-written News Redefines the Core Skills of Human Journalists." *Journalism Practice* 6 (5–6): 648–658.
- Van der Kaa, Hille, and Emiel Krahmer. 2014. "Journalist versus News Consumer: The Perceived Credibility of Machine Written News." Research Paper presented at the 2014 Computation + Journalism Symposium, Columbia University, New York. http://compute-cuj.org/cj-2014/cj2014_session4_paper2.pdf.
- Weischenberg, Siegfried, Maja Malik, and Armin Scholl. 2006. "Journalismus in Deutschland 2005. Zentrale Befunde der aktuellen Repräsentativbefragung deutscher Journalisten [Journalism in Germany 2005. Central Findings of a Representative Sruvey of German Journalists]." In *Media Perspektiven* 7: 346–361.
- Weischenberg, Siegfried, Maja Malik, and Armin Scholl. 2012. "Journalism in Germany in the 21st Century." In *The Global Journalist in the 21st Century*, edited by David H. Weaver and Lars Willnat, 205–219. New York: Routledge.
- Wolf, Cornelia. 2014. *Mobiler Journalismus. Angebote, Produktionsroutinen und Redaktionelle Strategien deutscher Print- und Rundfunkredaktionen* [Mobile Journalism. Services, Production Routines and Strategies of German Editorial Offices in Print and Audio]. Baden-Baden: Nomos.
- Young, Mary L., and Alfred Hermida. 2015. "From Mr. and Mrs. Outlier to Central Tendencies." *Digital Journalism* 3 (3): 381–397.
- Yu, Jin, Ehud Reiter, Jim Hunter, and Chris Mellish. 2007. "Choosing the Content Oftextual Summaries of Large Time-series Data Sets." *Natural Language Engineering* 13 (01): 25–49.

Konstantin Nicholas Dörr, Institute of Mass Communication and Media Research,
Media Change & Innovation Division, University of Zurich, Switzerland. E-mail:
k.doerr@ipmz.uzh.ch

Appendix A

List of Interviewees

Frank Feulner, CTO, Aexea Communication, Stuttgart, Germany.
Saim Alkan, CEO, Aexea Communication, Stuttgart, Germany.
Wolfgang Zehrt, CTO, Text-on, Berlin, Germany; now Textomatic AG.
Claude de Loupy, CEO and Co-founder, Syllabs (Data2Content), Paris, France.
Ehud Reiter, Chief Scientist, Arria NLG, Aberdeen, United Kingdom.
Johannes Sommer, Managing Director, Retresco, Berlin, Germany.
Johannes Bubenzner, Managing Director, 2txt NLG, Berlin, Germany.
Raul Valdes-Perez, CEO, OnlyBoth, Pittsburgh, United States.
Arden Manning, SVP Global Marketing, YSEOP, London, United Kingdom; United States, France.
John M. Pierre, CEO and Co-founder, Linguastat, San Francisco, United States.
Linda Hayes, CRO and COO, Linguastat, San Francisco, United States.
Edouard de Ménibus, Co-founder, Labsense, Paris, France.
James Kotecki, Media & Public Relations, Automated Insights, Durham, NC, United States.
Craig Booth, Data Engineer, Narrative Science, Chicago, United States.

Artikel III

When Reporters get Hands-On with Robo-Writing. Professionals Consider Automated Journalism's Capabilities and Consequences

Zusammenfassung

The availability of data feeds, the demand for news on digital devices, and advances in algorithms are helping to make automated journalism more prevalent. This article extends the literature on the subject by analysing professional journalists' experiences with, and opinions about, the technology. Uniquely, the participants were drawn from a range of news organizations –including the BBC, CNN, and Thomson Reuters—and had first-hand experience working with robo-writing software provided by one of the leading technology suppliers. The results reveal journalists' judgements on the limitations of automation, including the nature of its sources and the sensitivity of its «nose for news”. Nonetheless, journalists believe that automated journalism will become more common, increasing the depth, breadth, specificity, and immediacy of information available. While some news organizations and consumers may benefit, such changes raise ethical and societal issues and, counter-intuitively perhaps, may increase the need for skills – news judgement, curiosity, and skepticism – that human journalists embody.

Thurman, N., Dörr, K., & Kunert, J. 2017. When Reporters get Hands-On with Robo-Writing. Professionals consider Automated Journalism's Capabilities and Consequences. *Digital Journalism* [online first]. doi: [10.1080/21670811.2017.1289819](https://doi.org/10.1080/21670811.2017.1289819)

Copyright: <http://www.tandfonline.com>

The article is published as the version of record in Digital Journalism:
<http://www.tandfonline.com/doi/full/10.1080/21670811.2017.1289819>

WHEN REPORTERS GET HANDS-ON WITH ROBO-WRITING

Professionals consider automated journalism's capabilities and consequences

Neil Thurman, Konstantin Dörr and Jessica Kunert

The availability of data feeds, the demand for news on digital devices, and advances in algorithms are helping to make automated journalism more prevalent. This article extends the literature on the subject by analysing professional journalists' experiences with, and opinions about, the technology. Uniquely, the participants were drawn from a range of news organizations—including the BBC, CNN, and Thomson Reuters—and had first-hand experience working with robo-writing software provided by one of the leading technology suppliers. The results reveal journalists' judgements on the limitations of automation, including the nature of its sources and the sensitivity of its "nose for news". Nonetheless, journalists believe that automated journalism will become more common, increasing the depth, breadth, specificity, and immediacy of information available. While some news organizations and consumers may benefit, such changes raise ethical and societal issues and, counter-intuitively perhaps, may increase the need for skills—news judgement, curiosity, and scepticism—that human journalists embody.

KEYWORDS algorithmic journalism; automated journalism; computational journalism; journalism ethics; media economics; news production; professional skills; robot journalism

Introduction

In recent years, there has been increasing discussion about the use of computation in journalism practice (Anderson 2012; Örnebring 2010). With advances in algorithms and the availability of big data, the ways in which journalists search for, analyse, and distribute information are changing, giving rise to technical, cultural, economic, political, and ethical questions (Flew et al. 2012; Coddington 2015; Saurwein, Just, and Latzer 2015; Örnebring and Ferrer Conill 2016; Thurman 2011). These technical developments have also impacted on news writing. Using structured data, algorithms are generating news reports on crime statistics, sports matches, and company results. For example, the Associated Press uses natural language generation¹ technologies to automate up to 3700 quarterly earnings reports about US and Canadian companies (AI, n.d.). Until recently, software providers—like Arria, Narrative Science, AX Semantics, Retresco, and Automated Insights—have mostly developed bespoke products for their journalistic

clients. However, this business model is now being complemented by a self-service offer, which makes the technology available via Web-based interfaces, APIs (application programming interfaces), and plug-ins for tools such as Microsoft Excel. Potential users include other industries—such as e-commerce, financial services, and customer communication—as well as smaller organizations unable to afford custom-made software.

While most of the research into this so-called “automated journalism”² (Graefe 2016; Montal and Reich 2016) has focused on audience perceptions of automated content (e.g. Clerwall 2014; Graefe et al. 2016), on societal or economic implications (Dörr 2016; Latar 2015), or on ethics (Diakopoulos and Koliska 2016; Dörr and Hollnbuchner 2016; Montal and Reich 2016), there has also been some focus on the micro level, with the journalist as research object. van Dalen (2012) analysed journalistic reactions to the launch of StatSheet, a network of machine-written sport websites, by looking at reporting in 68 blog posts and newspaper articles. Carlson (2014) mimicked this method by examining how journalists wrote about text-generation software released by Narrative Science for use in journalism.

Although these studies provide useful analyses of the popular discourse around automated journalism, the journalists quoted had little or no practical experience using the technology. One exception was a study by Young and Hermida (2015) who interviewed seven employees of the *Los Angeles Times*’s Data Desk about the history, workings, and rationale behind its “The Homicide Report” website, which includes elements of automated reporting. Such studies are valuable for the insights they provide—from a professional, expert perspective—on issues such as the use of technology, changes to journalistic role conceptions, and labour and economic issues (Djerf-Pierre, Ghersetti, and Hedman 2016; Spyridou et al. 2013; Weischenberg, Malik, and Scholl 2006).

This study investigates the views of journalists on automated journalism by conducting workshops with professional practitioners. Ten British journalists from a range of media outlets and journalistic beats and with differing levels of responsibility and technical expertise were given the opportunity to work with software developed by one of the leading companies in natural language generation. The software, which the company made available for use in this study, provides various tools that enable journalists to create templates and thus automate news from structured data. The workshops were observed and semi-structured interviews were carried out with the participants.

In summary, the purpose of this qualitative, exploratory study is to elicit, report, and analyse the testimonies and opinions of a range of editors and journalists about automated journalism. These opinions were, in part, stimulated by giving the participants training and hands-on experience with software, from a leading provider, used to automate data-driven news texts. We follow Carlson’s (2014, 416) definition of automated journalism as “algorithmic processes that convert data into narrative news texts with limited to no human intervention beyond the initial programming”.

To help organize the presentation of this study’s results and the related contextual preamble, this article uses a framework consisting of three concepts: expertise, economics, and ethics.³

Literature Review

Expertise

The concept of expertise involves a consideration of professionals and their individual skills (Lewis and Westlund 2015, 454). Journalists are regularly required to revise their skills in the light of institutional, economic, and technological change (Willnat, Weaver, and Choi 2013). With the increased application of computing in journalism, certain skills, such as data analysis, programming, and visualization, have become more important (Mayer-Schönberger and Cukier 2013). Whereas traditional skills for investigative or interpretative reporting have long been part of journalism education (Hallin 2000), specific computational skills have been added to curricula relatively recently, if at all. It is, therefore, often the responsibility of employers or journalists themselves to provide or acquire the skills needed for contemporary news production (Deuze, Neuburger, and Paulussen 2004; Lewis and Usher 2014).

Some argue that journalists tend to accept rather than resist technologically driven change to job requirements and skills, seeing it as part of the natural evolution of their profession (Deuze 2007; Örnebring 2010; Creech and Mendelson 2015). However, others have highlighted the concerns that exist about the pressure to acquire new skills related to digitization and the introduction of new technologies (Willnat, Weaver, and Choi 2013).

Natural language generation software could lead to a fragmentation of work tasks, an increase in management control, and the deskilling (or upskilling) of workers (Örnebring 2010, 60; Rottwilm 2014, 13). Although early research has shown that the software may not yet be able to interpret or contextualize factual information fully (Dörr 2016; Graefe 2016), it could replace or augment some core journalistic skills such as accuracy and speed (Cleary and Cochie 2011, 79).

In light of the potential that natural language generation has to change the mix of skills required in the newsroom, and the uncertainty around journalists' reactions to its introduction, a number of related questions were included in the interviews.

Economics

Media economics has been defined as the study of "how media operators meet the informational and entertainment wants and needs of audiences, advertisers and society with available resources" (Picard 1989, 7). Media organizations, especially in journalism, are suffering from a shrinking advertising base, fragmenting audiences, and rising competition from mobile, social, and digital media (Anderson 2012; Picard 2010). Latzer et al. (2016) show that in the wider network economy the application of algorithms can have an economic impact in emerging, as well as existing, markets. They identify—amongst other applications—the economic potential of automated text generation, noting that algorithms can limit transaction costs for journalistic text production and lead to efficiency gains. Örnebring (2010) too writes of how technology "relieves" journalists of work that can be done by relatively inexpensive workers (including algorithms), an assessment shared by van Dalen (2012). A drive to reduce labour costs is not new to journalism, but with the increasing use of automation and algorithms in news production, it is acquiring a new dimension. While in earlier research

the economic implications of automated journalism were covered from a provider and market-driven perspective (Dörr 2016), this study evaluates the potentials and implications from the perspective of those who have first-hand professional experience of the settings in which the technology would work and in making the product it produces.

Ethics

Media ethics are a form of applied ethics focusing on the mass media system with its daily routines and professional practices (Horner 2013). Within media ethics, journalism ethics serve as a guide to approved societal moral values as well as helping to define journalists' work as professionals (Lewis and Westlund 2015, 459; Ward 2010).

Research shows that journalists feel digitization and automation complicate news production and make it more difficult to follow normative standards (Creech and Mendelson 2015, 149; Spyridou et al. 2013). The influence of algorithms and the increasing use of digital data raise new ethical questions, for example on:

- How journalists acquire, validate, and use digital data in news production (Bradshaw 2014; Zion and Craig 2014).
- The possibility of bias in the algorithms that power automation (Carlson 2014; Gillespie 2014; Shirky 2009; Thurman et al. 2016).
- Whether algorithms can "reason": draw appropriate conclusions in all contexts (Young and Hermida 2015).
- The transparency of code and data (Diakopoulos 2015; Diakopoulos and Koliska 2016; Thurman et al. 2016).

These questions have begun to be addressed in the literature. For example, Dörr and Holnbuchner (2016) identify, frame, and discuss ethical challenges around automated text production with regard to professional, organizational, and societal levels of responsibility in journalism. Complementing this conceptual approach, Montal and Reich (2016) analysed how 12 news websites bylined the automated news stories they published, concluding with a call for full disclosure.

To cover the ethical challenges of automated journalism from a professional perspective, questions on authorship, objectivity, and transparency were included in this study's interview framework and other ethical issues were discussed with respondents as they arose.

Methodology

Journalists' opinions were investigated qualitatively, as befitted this study's exploratory nature. Semi-structured interviews were conducted with 10 journalists who were given, in advance, training in, and an opportunity to work independently with, software that enables journalistic texts to be generated automatically. The journalists were observed by the researchers during these workshops, both to inform the conduct of the interviews and as a means of triangulating their responses.

Sample Selection

Automated news writing has found real-world application in a number of journalistic contexts. Beats such as sport, finance, and crime, where structured streams of data are readily available, and where reporting can be relatively formulaic, lend themselves to automation (Podolny 2015). Although automated news writing is not restricted to these beats, it was decided, when building the sample of journalists, to purposefully include journalists working in these areas. Journalists working in certain media (e.g. magazines and radio) and outside an institutional framework (e.g. freelancers) were excluded. Automated news writing is less likely to be implemented in these contexts, due, for example, to the types of journalism, and forms of presentation, utilized in radio and magazine journalism, and to the resource constraints under which freelancers operate. Journalists were identified using a commercial database and a total of 641 journalists were selected to receive an email invitation to participate. Because both the workshops and interviews were to be conducted in London, only journalists from Greater London were contacted.

A total of 22 journalists expressed an interest in participating. The final sample of 10 journalists was constructed purposefully in order that a variety of beats, levels of seniority, technical expertise, and a mixture of those working for private and public institutions were represented (see Table 1). Having journalists at different levels of seniority was considered important because of the divergent opinions they might exhibit. For example, senior managers may view the technology positively because of its potential for reducing labour costs and for increasing control over employees and processes. Conversely, rank-and-file journalists might consider the technology as a threat to their jobs. Journalists from the BBC were purposefully included in the sample, not only because the BBC is the United Kingdom’s largest news organization, but also because journalists from such a well-resourced public service broadcaster might be in a better position to experiment with new technologies because of its relatively secure funding model. A range of technical expertise was sought because the technology might, simultaneously, deskill some, while providing the opportunity for others to be

TABLE 1
The sample of journalists used in this study, showing their beat, employer, and professional position

Code	Beat	Media outlet	Position
A	Sports	CNN	Sports journalist in executive position
B	Sports	Thomson Reuters Television	Sports journalist in executive position
C	Sports	Thomson Reuters	Sports journalist in executive position
D	Data	Trinity Mirror	Data journalist
E	Data	BBC	Data journalist
F	General news	BBC	News reporter/journalism trainer
G	Crime	<i>The Sun</i>	News reporter
H	Finance	Thomson Reuters	Financial journalist in executive position
I	Finance	Thomson Reuters	Financial journalist, correspondent
J	Finance	BBC	News reporter, working in research and development

upskilled, with corresponding differences in opinions. With the exception of Journalist E, none of the participants had experimented with this kind of software before.

Workshop Content and Format

The software workshops and interviews were held on 12 and 13 May 2016 at the Interaction Lab of City, University of London. Each workshop accommodated a maximum of three journalists who worked on their own PC, learning how to use the software and, ultimately, creating a template that could produce news reports in an automated fashion. Each of the four workshops lasted approximately two and a half hours, and was led by a native English speaker (one of the authors of this article).

The software used in the workshop is an online application for generating journalistic texts from structured data. Access was provided by one of the leading companies in the field who service a variety of large journalistic clients. In order that an appropriate level of technical instruction could be given, the researchers were trained by the company twice and pre-tested the workshop format with a group of students. Due to non-disclosure and anonymity agreements, neither the company, the software platform, nor the journalists can be identified.

To use the software the journalists uploaded structured data. Next, they constructed a template using a graphical user interface that utilizes the variables in the data to generate the news texts. At a simple level, variables from the data file can be inserted directly. For example:

[Westminster] recorded the highest number of offences for any London borough with [3,899].

Where [Westminster] and [3,899] are the variables that are replaced in each new story produced by the template. The formatting of variables can be modified, e.g. their capitalization, the number of decimal places, and so on. If the data do not contain all the variables a journalist might wish to use as part of the template, new ones can be created by transforming existing variables using simple mathematical functions. In order to add variability to the final outputs, individual words in the template can be assigned an unlimited number of synonyms that are picked at random when the texts are generated. Finally, branching offers the ability for words, sentences, or whole paragraphs to be inserted depending on certain conditions. For example, if one variable is equal to, greater than, or less than another. Data variables and synonyms can be used within branches and branches themselves can contain branches, making it possible to create templates of great complexity.

In the workshops, the journalists were grouped according to their beat. The data-sets used took inspiration from examples provided by the company and were collated by the researchers. The first data-set, for the finance journalists, used employment data on the United Kingdom released by the Office for National Statistics. The second, for the general news reporters, consisted of crime statistics relating to London boroughs released by London's Metropolitan Police. The third, for the sports journalists, consisted of data relating to several English Premier League soccer matches.

The employment data-set consisted of 22 separate variables, such as: "non UK nationals in employment". The crime data-set had more variables (52), such as: "total

drug offences 11 month average". The sports data-set consisted of 61 variables, such as: "scorer of first home team goal". Some of the variables were taken directly from the original data sources, and others were the result of transformations made to the data that would be useful to journalists, e.g. so that they could make comparisons across time periods or see outlying values.

Data Collection and Analysis

The journalists were observed during the workshops by three researchers (including two of this article's authors) from behind a one-way mirrored wall. Although the journalists were aware of this set-up beforehand, they could not see or interact with the observers during the workshops. As well as being able to see the journalists' body language through the mirror, the observers had an audio feed of proceedings and could see how the journalists were interacting with the software via screens that echoed the output of each participant's computer monitor. Notes were taken during the observations.

Immediately following the completion of each workshop, participants were interviewed by the observers. Each interview lasted between 37 and 60 minutes and was digitally recorded and transcribed verbatim. The coding and analysis followed a fairly standard process, such as described by Weiss (1994, 151–182) and Creswell (1994), of iterative issue identification, mark-up, and theme development. Quotations used in this article have been edited for concision and clarity, while preserving the meaning. The questions that made up the semi-structured interview plan were written following a literature review and grouped into overarching themes that, broadly, fitted within this article's conceptual framework of expertise, economics, and ethics.

Results

Expertise

There was a significant amount of comment on the constraints imposed by the technology, in particular its reliance on single, isolated data streams and the necessity to predict news angles (or "top lines"⁴) in advance. Journalists felt these constraints meant that items produced in this way would lack the context, complexity, and creativity of traditional reporting.

Sourcing. As mentioned, journalists believed that one of the technology's main constraints, even in areas such as sports, crime, and finance reporting, where automation has been considered to have the most potential, was its reliance on single, isolated data streams. A data feed⁵ about a single English Premier League soccer game of the type that could be used to power an automated match report can contain "2,000 detailed events", from the result down to very granular data (Opta, n.d.). Those data, however, relate to what happened on the pitch, which, a senior sports journalist at CNN (A) said, might be "largely irrelevant" to the story a journalist covering a fixture might want to write. Other "big issues" off the pitch could be more important, he said, giving the example of a game between West Ham United and Manchester United in

May 2016. In that match the “riot before the game”, the fact that by losing Manchester United were unlikely to be able to qualify for the Champions League, and the fixture’s place in history as the last West Ham would play at their Upton Park ground after 112 years were more important than the events that took place on the turf. These contextual factors, the journalist thought, “can’t be programmed in”, meaning that the data might only present “10 per cent of the story”.

A journalist at Thomson Reuters (I) made a similar point, giving the example of a financial story based around a data release from the European Central Bank (ECB). While a data feed might communicate that the ECB had kept interest rates unchanged, it could not, he said, convey the “semantics” of the announcement, e.g. how the ECB’s president “answers questions”. Furthermore, “a lot of the really important stuff”, he said, “is being briefed off-the-record by bankers to journalists”.

Other respondents (A, F, and B) also mentioned that using multiple sources is a characteristic of good journalism and that such sourcing, e.g. from “the front line of a war zone”, often has to be done in the analogue world, by a real human being (H).

Interrogating the data. The workshop provided an illustration, for some participants, of how, even within a single data-set, it can take “a human” (G) to spot irregularities that could have important implications for how the story is told. One of the data-sets used in the workshop, scraped from the website of the Metropolitan Police, was on crimes in London. It showed total crimes for February 2016 were 58,175 (down from 60,269 in January), a figure that, on the surface, “looks much better” than the previous month, said Journalist C. An automated, data-driven news item based on these numbers might have chosen to highlight this fall. However, as both C and G spotted, February had two fewer days than January, meaning the average number of crimes per day in February was actually higher.

Both journalists said they would have used this anomaly as a starting point for further enquiries, with the crime reporter at *The Sun* (G) suggesting that a failure to mention the daily average in any data release might be indicative of an attempt by the police to “manipulate the figures”. This ability to interrogate is, thought a news reporter/trainer at the BBC (F), what “any good journalist will do”. Even the data journalist at Trinity Mirror (D) agreed that “there’s always going to be things within data feeds that will stand out more to a human than it will to a computer”.

The human angle. The lack of a human angle in the texts generated by data-driven automation was a concern for some. One journalist (G) said that even when reporting on companies—a high-profile use of this type of technology (AI, [n.d.](#))—the City desk at his newspaper, *The Sun*, would “always look for the human angle”, an approach he did not believe was confined to tabloids. He gave the example of the collapse of the British retailer BHS. Although there was a “numbers” side to the story, people wanted to know about the “flamboyant lifestyle and arrogance” of the former owner, Sir Philip Green, and the effects on the company staff and pensioners of the “awful way the company was run”. Unless stories are about humans “they are pretty dull”, he said. Other journalists (C, J, and F) agreed, with one (F) saying that “journalism is about telling stories” that involve human beings whose lives are not easily quantifiable into programmable data.

Known unknowns. The requirement to template stories in advance was seen as a drawback. A BBC journalist (J) realized, after using the software, “how limiting writing everything in advance is. You can’t get a reaction to those numbers, you can’t explain or interrogate them, because you wrote it all before the numbers came out”. His colleague (F) at the BBC agreed: “news comes out of the blue sometimes. I am not sure if you can write up in advance something you didn’t expect to happen”. Predictably given his technical expertise, the data journalist at Trinity Mirror (D) was more confident that, given sufficient development, templates could account for multiple outcomes, however, this would still be “very, very complicated” to achieve, he said.

Playing with the data. Traditional data journalism is often about using data to produce a single story, sometimes from multiple sources.⁶ The software the participants used in the workshop is different. Although also data-driven, it exists to output multiple news items from a single source. Some journalists expressed frustration with the software’s functionality, perhaps because they were evaluating it against the more investigative styles of data journalism. For example, J found it “quite frustrating” because he wanted “to be able to play around with these data-sets much more, looking for the story in the numbers, not just take the numbers and turn them into a templated story”.

This reporter summarized the limitations of automated news text generation by contrasting it with the multi-sourced, interrogated, and contextualized journalism humans can produce: “Journalism is about finding out stuff that other people are not able to find out about. It is about asking difficult questions. It is about putting things together to understand what is really going on”.

Perhaps unsurprisingly, when evaluating the output generated in the workshop by the software, some participants were less than impressed. A senior sports journalist at CNN (A) said that a fully manually written story would have been “much better”. He thought the output was “throw away, repetitive, not particularly interesting. There is no nuance or creativity”. J agreed, saying: “I would never, ever, ever have written a story like that. Where is the analysis, context, and quotes from participants or experts?”

Despite the limitations identified by the journalists, respondents did believe the technology could have a role in augmenting their work.

Opening narrative seams. One of the journalists who thought that automation might have a role in helping spot stories in data, C, gave the example of a soccer data feed which showed that Leicester City could win a match with just 30 per cent ball possession. This had, he said, “opened up a rich seam to be mined for a narrative”, prompting questions about how a team could be successful with far less possession than “accepted wisdom” suggests is required. A data journalist at Trinity Mirror (D) agreed that the technology might be useful as a way of “creating internal briefing notes” from data feeds to highlight information that journalists might want to follow up, sentiments shared by journalists F and G.

There were conflicting opinions about whether the technology would, in the long run, help journalists cope with the increasing amount of data being released. A data journalist at Trinity Mirror (D) thought that “software could make data more intuitive for people”. On the other hand, a BBC reporter (J) felt that this particular approach, using technology to turn numbers into words, might undermine efforts to make journalists more numerate.

Pleasure and propensities. For several journalists, the experience of creating news items this way was difficult, irritating, and did not utilize their innate abilities. A senior sports journalist at CNN (A) “didn’t find [the software] that intuitive to use” and the process did not provide him with the occupational rewards he was accustomed to: “What gives me pleasure as a journalist is the use of language and creativity and [this software] is the complete opposite”. A BBC reporter (J) went further: “I can’t tell you how bored I was. It is essentially computer programming and I hate computer programming”. Other journalists (C and F) also alluded to a mismatch between the skills required to produce data-driven news and the skills journalists often identified within themselves. Journalist F did acknowledge, however, that resistance could, in part, be “generational”. He thought that for “digital natives” the “idea of a software program writing a story” is not as “outrageous” or “frightening” as it is for older journalists. Of course, in most real-world implementations of the technology, journalists would be unlikely to develop automation processes without technical assistance.

Complementarity. A number of journalists discussed the technology as complementary to their work. Two senior sports journalists at Thomson Reuters (B and C) suggested that in “the first stage of the news cycle, the straightforward facts could be automated”, with the human journalist coming in “further up the value chain” to focus on what “warrants further exploration” (C). These sentiments were echoed by journalists G, I, and H. The data journalist at Trinity Mirror (D) thought that, because of the current limitations of the technology, it may be better for software to produce “the body of the article” and then for “the introductions” to be written by a human afterwards in order to “pull out the top line”.

Economics

A significant number of the interviewees believed that automated news writing had the potential to reduce costs, increase speed, and serve more clients with more individualized content, advantages that have also been identified in popular and academic discourse around the technology (Carlson 2014; Dörr 2016). Some of the respondents in executive positions were aware of projects currently under way in their news organizations to roll out automation, including in the production of multimedia content. However, journalists believed that the technology, in its current form, only had application in limited contexts and was unlikely to fundamentally change the economic dynamics of the industry.

Reducing costs. Journalists at CNN and Thomson Reuters thought that the technology offered an opportunity to “reduce costs” (A) by replacing “expensive staff” (C) who were doing “fairly simplistic and time-consuming work” (C). They believed there was an inevitability to these developments, in part because of the “[financial] conditions in the industry currently” (C). Although, as will be shown later, journalists believed application of the technology was limited, it was considered that there was a market for what it could produce: “Some customers are happy with an unsophisticated product. And if we can derive that instantly at a low cost, then clearly that’s the way we are going to go”, said a senior sports journalist at Thomson Reuters (C).

Increasing speed. Journalists (C, H, and B) saw the technology as having potential to increase the speed with which they could deliver material to their audience. For example, a senior sports journalist at Thomson Reuters (C) saw the question of whether to use automation as a “speed decision”, explaining that, as a global news agency, they had to be “very fast and accurate and automating is one way to improve both those things”.

Expanding coverage. Journalists at a range of news organizations saw the technology as having potential to expand the volume of material produced and to cater to different devices—such as mobile phones—different geographical markets, and individualized content needs.

At least three journalists thought that automated news writing might have utility in the production of content for mobile platforms. A senior sports journalist at Thomson Reuters (C) explained that: “People don’t want to read 5000 words on their phone. They want a really nice picture and two or three paragraphs on what’s happened. This is what we’re thinking currently about our automated future”.

This same journalist saw potential to tailor the text of a story for different markets: “You could automatically generate a story about, for example, Leicester versus Liverpool [soccer match] and send a different version of the story to Liverpool and Leicester, and have a third one for neutrals”.

A data journalist at Trinity Mirror (D) envisaged addressing different demographics rather than geographical markets, using different styles of language in order, for example, to version the same story for tabloid and middle-market readers.

Other journalists thought that it would be possible to use the technology to produce textual content that was even more personalized. A BBC journalist (J) thought that stories personalized to “a reader’s street or neighbourhood would be of potential use. It’s about re-versioning, a big thing at the BBC and in the industry. How do you re-version with no money and give a more personalized service?”

The BBC data journalist (E) thought that the technology might even be able to write stories that provided “context and background” appropriate to the level of information an individual needed, with reference to, for example, what the person had already read.

A journalist (H) at Thomson Reuters had the most concrete ideas for how automation could expand coverage: “It gives us an opportunity to grab massive amounts of data and package it up in a way our clients can use” Packaging that data might mean, he suggested, taking “economic data released in Turkey in Turkish” and automatically writing a story in English incorporating that data. The same principle could apply, he said, for any other combination of the “50 languages” in which Reuters operates.

Roll out. The opportunities for reducing costs, increasing speed, and expanding coverage that the respondents identified are compelling enough for projects to already have got under way in at least two of the news organizations represented by our interviewees. A senior sports journalist at Thomson Reuters (C) said that “we are working with an American company on a project to cover US sports, something we wouldn’t have the resources to cover manually”. For “big sports” like the US NBA basketball league, a product was being developed “mainly for mobile devices”. With this

product, produced in partnership with a US-based technology provider, slideshow-style videos would be created automatically from photos with a narration “extracted” from an “automated three-paragraph story”. Thomson Reuters’ developments in automation are not, however, limited to sports reporting. A financial journalist at Thomson Reuters (I) said that already “quite a lot of ‘snaps’ [short news alerts that are sent out before a full story is published] are automated”. Another financial journalist (H) at the firm confirmed that they were “looking extensively” at automation in the finance sector; indeed, the senior sports journalist (C) said, “we’re looking at it in all parts of the company”.

Although projects are under way at Reuters, representatives of other news organizations were less enthusiastic about immediate adoption. A senior sports journalist at CNN (A) said that “from what I saw I wouldn’t want to” adopt the technology, but rather “use that money to employ a person”. These sentiments were shared by some journalists from the BBC. One (F) did not think it was “scalable to the BBC”, and his colleague (J), working in BBC research and development, left the workshop “less impressed than I was before today”. He continued: “Before the workshop I was wondering whether we should put some effort in researching on robot journalism, and after today my conclusion might be ‘No’”. As it happened, however, one of the other participants from the BBC, a data journalist (E), had already been involved in at least one project, seemingly off the radar of the other BBC journalists in our sample, which utilized automation. The example given involved automatically producing short updates on sports results as part of the BBC’s online coverage of the London 2012 Olympics.

Limited potential. Although some journalists spoke about the economic opportunities and current plans for automation, there was, for the most part, consensus that applications were limited at the current time. Sports and finance were the most frequently mentioned areas of application (F, B, and G), with weather, crime (F), election results (G), and “anything where there is a big data-set” (F) also suggested. Even the BBC data journalist acknowledged that “there is lots of reporting that you can’t really template out”. For the crime reporter at *The Sun* (G), the limitations were as much to do with the style that automation imposes as they were to do with the requirements for structured data. He considered the output to be “too bland for our readers” who wanted a “ribald” and “hard-edged” approach that plays with language.

There were wider concerns that the technology was adding little value and was unlikely to increase the appeal of news outlets in a way that would make any significant difference to their audience share. A sports journalist with Thomson Reuters Television (B) said, “I don’t know how much value is in this kind of content. Is it attractive to consumers? Not really. It is boring”. The CNN journalist agreed, adding that “people want something a little bit more creative”, such as, a BBC reporter (F) suggested, “stories about human beings, with proper top lines, written by proper journalists who know what a story is”.

Although the software was considered a “niche product” (H) with limited potential, some journalists acknowledged that automation was still in its infancy and “nobody knows if it will be able to cover more topics in the future” (H). Indeed, there was a feeling among most respondents that automation would become more common.

Ethics

Participants discussed a range of ethical issues concerning automated news writing, including: transparency, bias and balance, verification, as well as wider societal concerns.

Transparency. Respondents held mixed views about whether and how automated news should be labelled, indicative, perhaps, of the complex history of authorship and bylining in journalism (see, e.g., Reich 2010) and the novelty of thinking about such issues in the context of non-human actors. Overall, however, a majority came out strongly or conditionally in favour of transparency. For example, a senior Thomson Reuters journalist (C) thought that there was an obligation to be transparent, in part because audiences were likely to be unaware that a story might be “data-driven”, but also because of normative expectations that information would be provided about a news item’s provenance.

The form that such transparency might take prompted discussion about whether a simple byline—e.g. “This story was written by a computer”—was sufficient. The BBC data journalist (E) thought not, proposing that a fuller “methodology” be posted. Another BBC journalist (F) agreed, saying that news organizations should “show [their] workings and give people as much information as they can handle” in order that readers could be informed to make decisions on partiality.

However, the nature of automated, data-driven news, relying, as it does, on decisions made, and data generated, by both humans and software routines, raises questions about what the “methodology” or “workings” would reveal. A BBC news reporter (J) talked about the complexity of determining authorship when “bots are part human and part bot”, even going further by calling into question the extent to which automated news was, in fact, autonomous: “it is written by a human being in the first place”. A crime reporter from *The Sun* (G) and a Thomson Reuters journalist (I) also discussed how the level of technological involvement in the production of a news item would influence how an item was labelled, with only something “entirely written by computer” (G) needing to be labelled as such.

Despite broad consensus on transparency and labelling, two of the respondents did not think that it was necessary to flag to the audience that a piece had not been written by a human. One, a data journalist with Trinity Mirror (D), justified this view using similar reasoning to the BBC news reporter (J), saying that “at some level, a journalist has built the structure for the story”.

Bias and balance. Three journalists suggested that processes of automation and “datafication” had the potential to reduce bias in reporting. One, from Thomson Reuters Television (B), acknowledged “bias” in “human journalism” and suggested that “there is a lack of bias in technology”. A reporter from *The Sun* (G) also talked about “manipulation” in human journalism and thought that so-called “robo-news” could “present the facts and the figures as they are”, which “could be quite a good tool for democracy”. Similar sentiments were espoused by a BBC journalist (F) who contrasted what he believed to be the “hatchet job” the *Daily Mail* had been doing on Jeremy Corbyn, the leader of Britain’s Labour Party, “every day” with the “black and white” of data-driven news, given, of course, that the “data is complete and accurate”.

However, another BBC News journalist (J) took a different stance, concerned about the possibility that the volume of media content it is possible to produce through automation could mean that the “prejudices” of a single individual or organization, encoded into an algorithm, were widely disseminated. He suggested this could happen either via news items published by the mainstream media, or through social media posts directly consumed by the audience or used by journalists as a source: “More and more news organizations are following what is trending on social media. If someone can get enough stuff out, they can influence the news agenda and get their views into the mainstream”.

Accuracy and verification. Two Thomson Reuters journalists (C and H) felt that automation had the potential to “eliminate human error” (C), in particular because of the speed with which news items had to be produced and published within a news agency context, although they did acknowledge that “automation is only as good as the data” (H).

Although an algorithm might be less likely to make a simple mechanical error than a human reporter, there was concern that automation would make verification less likely and oversight harder. A CNN journalist (A) questioned whether a computer program could do what journalists do and “test whether the information was true”, a sentiment echoed by a Thomas Reuters Television journalist (B) who said that because the data may be “deliberately corrupted”, the “human instinct” for detecting the “many glitches” that could happen was still required. Indeed, he went as far as to say that he did not believe that the output of automated news “could be trusted to go out unchecked at the moment”. The two data journalists in the sample were, predictably, more confident that, albeit with “editorial oversight”, “validation rules”, and trust in the data source (E), automation did not present “too much of a problem” (D) in terms of verification.

Personal and societal consequences. The journalists’ ethical concerns were mostly focused at the level of journalistic routines; however, one (F), from the BBC, expressed concern that the technology would contribute to information overload, making it harder for journalists and the public to pick out what was important from the sea of information:

What this software does is put more information out there. Is that a good thing? Neo-liberal theory would say yes, because we all want as much choice as possible. But do we? Sometimes it is very difficult to work out whether something is true. It is the job of a journalist to be the filter and get a simple story out that people can understand.

Conclusion

This article extends the growing body of literature on automated journalism by eliciting and analysing professional journalists’ and editors’ experiences with, and opinions about, the technology. Uniquely, the participants were drawn from a range of news organizations and had—at the very least—first-hand experience working with robo-writing software from one of the leading suppliers.

The study provides insights into how some large news and information providers now use automation, and initiatives we can expect to appear in the near future.

For Thomson Reuters, those initiatives include using automation to cover beats it would otherwise be uneconomical to cover, such as some US sports, and utilizing the technology to produce not only text but also short audiovisual reports. At the other news organizations represented in this study, where there is less emphasis on high volumes of fast, factual reporting and more emphasis on analysis and entertainment, automation was far less evident.

Beyond such factual findings, this study's key contributions are in its documentation and analysis of journalists' judgements on the potentials and limitations of automation in the context of the variety of roles they perform and outputs they produce, and on the wider ethical and societal implications. Six main limitations of automated journalism were identified:

1. Its reliance on single, isolated data streams.
2. The one-dimensional nature of the quantitative data feeds it relies on.
3. The difficulties of interrogating that data.
4. The lack of human angles in the texts generated.
5. The requirement to template stories—predicting “top lines”—in advance.
6. The difficulty of working creatively with the data in the templating process.

As the participants recognized, the technology is still at a relatively early stage of development and some of these limitations may, ultimately, be moderated. For example, the “Chief Visionary Officer” of one service provider, AX Semantics, explained that:

- their technology includes “a meta-database, called ‘World Knowledge store’”, which can be utilized to provide additional context to automated stories;
- given sufficient technical expertise, it was possible to “draw complex conclusions” from the data and modify it in various ways;
- they provided functionality to detect whether “data is well-shaped”, “missing crucial information”, or “exceeds certain thresholds”; and
- extra dimensions (such as “image recognition”) are likely to be added “within the next year” (Frank Feulner, personal communication, 27 October 2016).

However, even such developments, and others that may follow, are unlikely to overcome what we believe to be some fundamental limitations with automated journalism. For example, secondary “meta-level” data streams may be able to provide historical background and static facts but are less likely to be able to provide the contemporaneous context that is essential to much reporting. Even when systems have, as Feulner predicts, the ability to “analyse sentiment and soft facts” around topics and “identify irregular or deviant data points”, they are unlikely to be able to understand all the nuances of human expression that help determine how events are reported or consistently recognize the most important news angle, the “top line”.

This said, it seems clear that automated journalism will be used increasingly to produce simple factual reports, to increase the speed with which such reports are published, to cover topics currently below the threshold of reportability (such as minority/minor league sports), to re-version or adapt content for specific devices (such as mobile phones) or individual tastes, and to help journalists spot stories in data.

Such a future may bring financial benefits to news organizations although not, in the short term, to the workers who are likely to be replaced.⁷ Our respondents provided some insights into the ethical questions that arise from such changes in news

production. While recognizing automation's potential to reduce bias and inaccuracy in human reporting, they also saw challenges in maintaining standards of verification and balance, and deciding who, or what, should be credited and held accountable for the output.

The consequences for society more widely are also double-edged. On the one hand, some consumers may extract value from the depth, breadth, specificity, accessibility, and immediacy of the information produced by automated journalism. For others, however, such changes are likely to increase the difficulties of navigating a world already saturated with information. In the context of news and information, then, automation may actually increase the need for the very human skills that good journalists embody—news judgement, curiosity, and scepticism—in order that we can all continue to be informed, succinctly, comprehensively, and accurately, about the world around us.

Limitations

The results of this exploratory study should, of course, be interpreted in the light of its constraints. Firstly, the participants, although sampled purposefully to include a range of ages, skills, and professional experience, do not constitute a representative cross-section of journalists. Nevertheless, their number, 10, exceeds McCracken's (1988, 17) guidance that for "many research projects" based on qualitative semi-structured interviews, "eight respondents will be perfectly sufficient". Secondly, the software used by the participants, although provided by a technology company that is one of the leaders in the field and has high-profile journalistic clients, is one of a number of platforms available. The participants' opinions on the potentials and limitations of automated journalism have, therefore, been coloured by their experience of this particular platform. Thirdly, although the participants were introduced to the platform in small groups (a maximum of three), the limited time (150–180 minutes) they spent using the software meant that those who were less technically proficient had only a limited opportunity to explore, independently, the software's capabilities. The time was sufficient, however, for several of the participants to be able to iterate at least one story template that utilized a high proportion of the available functionality.

ACKNOWLEDGEMENTS

The authors would like to thank Hedvig Sandbu and Stuart Scott for their assistance and "the company" for providing access to and training on the natural language generation software platform.

DISCLOSURE STATEMENT

No potential conflict of interest was reported by the authors.

FUNDING

The research was supported by a research grant from the Volkswagen Foundation [grant number A110823/88171].

NOTES

1. Natural language generation is defined as software and computer systems which automatically produce human (natural) language from a computational representation of information (Reiter and Dale 2000).
2. The terms “algorithmic journalism” (van Dalen 2012; Dörr 2016) and “robot journalism” (Carlson 2014) are also used to describe the phenomenon.
3. This framework takes inspiration from Lewis and Westlund’s (2015) article on big data in journalism.
4. Defined by *The Oxford English Dictionary* as “the headline of a newspaper”, but used more widely by journalists when referring to the most newsworthy aspect of a story.
5. Such as those provided by Opta, the “Official Media Data Partner of the Premier League”, <http://www.optasports.com/events/premier-league.aspx>.
6. For example, the award-winning “What if the Syrian Civil War Happened in Your Country?”, by Public Radio International, that uses data from the United Nations, the Syrian Observatory for Human Rights, and the World Health Organization.
7. For example, one of our respondents (I) identified the work being done by the 400 Thomson Reuters employees in Bangalore—such as producing reports on “share price moves in the morning”—as “very, very easy” to automate.

REFERENCES

- Al. n.d. “The Future of Finance is Automated.” <https://automatedinsights.com/ap>
- Anderson, Chris W. 2012. “Towards a Sociology of Computational and Algorithmic Journalism.” *New Media and Society* 7 (15): 1005–1021.
- Bradshaw, Paul. 2014. “Data Journalism.” In *Ethics for Digital Journalists: Emerging Best Practices*, edited by Lawrie Zion and David Craig, 202–219. New York: Routledge.
- Carlson, Matt. 2014. “The Robotic Reporter: Automated Journalism and the Redefinition of Labor, Compositional Forms, and Journalistic Authority.” *Digital Journalism* 3 (3): 416–431.
- Cleary, Johanna. and Meredith Cochie. 2011. “Core Skill Set Remains Same in Newspaper Job Ads.” *Newspaper Research Journal* 32 (4): 68–82.
- Clerwall, Christer. 2014. “Enter the Robot Journalist: Users’ Perceptions of Automated Content.” *Journalism Practice* 8 (5): 519–531.
- Coddington, Mark. 2015. “Clarifying Journalism’s Quantitative Turn: A Typology for Evaluating Data Journalism, Computational Journalism, and Computer-Assisted Reporting.” *Digital Journalism* 3 (3): 331–348.

- Creech, Brian, and Andrew L. Mendelson. 2015. "Imagining the Journalist of the Future: Technological Visions of Journalism Education and Newswork." *Communication Review* 18 (2): 142–165.
- Creswell, John W. 1994. *Research Design: Qualitative and Quantitative Approaches*. Thousand Oaks, CA: Sage.
- van Dalen, Arjen. 2012. "The Algorithms behind the Headlines: How Machine-Written News Redefines the Core Skills of Human Journalists." *Journalism Practice* 6 (5–6): 648–658.
- Deuze, Mark. 2007. *Media Work*. Cambridge: Polity Press.
- Deuze, Mark, Christoph Neuberger, and Steve Paulussen. 2004. "Journalism Education and Online Journalists in Belgium, Germany, and the Netherlands." *Journalism Studies* 5 (1): 19–29.
- Diakopoulos, Nicholas. 2015. "Algorithmic Accountability: Journalistic Investigation of Computational Power Structures." *Digital Journalism* 3 (3): 398–415.
- Diakopoulos, Nicholas, and Michael Koliska. 2016. "Algorithmic Transparency in the News Media." *Digital Journalism*. doi:10.1080/21670811.2016.1208053.
- Djerf-Pierre, Monika, Marina Gherse, and Ulrika Hedman. 2016. "Appropriating Social Media: The Changing Uses of Social Media among Journalists across Time." *Digital Journalism* 4 (7): 849–860. doi:10.1080/21670811.2016.1152557.
- Dörr, Konstantin N. 2016. "Mapping the Field of Algorithmic Journalism." *Digital Journalism* 4 (6): 700–722.
- Dörr, Konstantin N., and Katharina Hollnbuchner. 2016. "Ethical Challenges of Algorithmic Journalism." *Digital Journalism*. doi:10.1080/21670811.2016.1167612.
- Flew, Terry, Christina Spurgeon, Anna Daniel, and Adam Swift. 2012. "The Promise of Computational Journalism." *Journalism Practice* 6 (2): 157–171.
- Gillespie, Tarleton. 2014. "The Relevance of Algorithms." In *Media Technologies: Essays on Communication, Materiality, and Society*, edited by Tarleton Gillespie, Pablo Boczkowski and Kirsten Foot, 167–194. London: MIT Press.
- Graefe, Andreas. 2016. "Guide to Automated Journalism." <https://towcenter.gitbooks.io/guide-to-automated-journalism/content/>
- Graefe, Andreas, Mario Haim, Bastian Haarmann, and Hans-Bernd Brosius. 2016. "Readers' Perception of Computer-Generated News: Credibility, Expertise, and Readability." *Journalism*. doi:10.1177/1464884916641269.
- Hallin, Daniel. 2000. "Commercialism and the Professionalism in the American News Media." In *Mass Media and Society*, edited by James Curran and Michael Gurevitch, 218–237. London: Arnold.
- Horner, David S. 2013. *Understanding Media Ethics*. Brighton: SAGE.
- Latar, Noam Lemelshtrich. 2015. "The Robot Journalist in the Age of Social Physics: The End of Human Journalism?" In *The New World of Transitioned Media: Digital Realignment and Industry Transformation*, edited by Gali Einav, 65–80. Wiesbaden: Springer.
- Latzer, Michael, Katharina Hollnbuchner, Natascha Just, and Florian Saurwein. 2016. "The Economics of Algorithmic Selection on the Internet." In *Handbook on the Economics of the Internet*, edited by Johannes Bauer and Michael Latzer, 395–425. Cheltenham: Edward Elgar.
- Lewis, Seth C., and Nikki Usher. 2014. "Code, Collaboration, and the Future of Journalism: A Case Study of the Hacks/Hackers Global Network." *Digital Journalism* 2 (3): 383–393.
- Lewis, Seth C., and Oscar Westlund. 2015. "Big Data and Journalism." *Digital Journalism* 3 (3): 447–466.

- Mayer-Schönberger, Viktor, and Kenneth Cukier. 2013. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. London: John Murray Publishers.
- McCracken, Grant. 1988. *The Long Interview*. Thousand Oaks, CA: Sage Publications.
- Montal, Tal, and Zvi Reich. 2016. "I, Robot. You, Journalist. Who is the Author?." *Digital Journalism*. doi:10.1080/21670811.2016.1209083.
- Opta. n.d. "Covering the World of Football in the Greatest Detail." <http://www.optasports.com/media/836304/final-opta-coverage.pdf>
- Örnebring, Henrik. 2010. "Technology and Journalism-as-Labour: Historical Perspectives." *Journalism* 11 (1): 57–74.
- Örnebring, Henrik, and Raul Ferrer Conill. 2016. "Outsourcing Newswork." In *The SAGE Handbook of Digital Journalism*, edited by Tamara Witschge, Chris W. Anderson, David Domingo and Alfred Hermida, 207–221. London: Sage.
- Picard, Robert G. 1989. *Media Economics: Concepts and Issues*. Newbury Park, Calif.: Sage.
- Picard, Robert G. 2010. "A Business Perspective on Challenges Facing Journalism." In *The Changing Business of Journalism and Its Implications for Democracy*, edited by David Levy, and Rasmus Kleis Nielsen, 17–24. Oxford: Reuters Institute for the Study of Journalism, University of Oxford.
- Podolny, Shelley. 2015. "If an Algorithm Wrote This, How Would You Even Know?." *New York times*, 7 March. http://www.nytimes.com/2015/03/08/opinion/sunday/if-an-algorithm-wrote-this-how-would-you-even-know.html?_r=0.
- Reich, Zvi. 2010. "Constrained Authors: Bylines and Authorship in News Reporting." *Journalism* 11 (6): 707–725.
- Reiter, Ehud, and Robert Dale. 2000. *Building Natural Language Generation Systems*. Cambridge: Cambridge University Press.
- Rottwilm, Philipp. 2014. *The Future of Journalistic Work: Its Changing Nature and Implications*. Oxford: Reuters Institute for the Study of Journalism.
- Saurwein, Florian, Natascha Just, and Michael Latzer. 2015. "Governance of Algorithms: Options and Limitations." *Info* 17 (6): 35–49.
- Shirky, Clay. 2009. "A Speculative Post on the Idea of Algorithmic Authority." <http://www.shirky.com/weblog/2009/11/a-speculative-post-on-the-idea-of-algorithmic-authority/>
- Spyridou, Lia Paschalia, Maria Matsiola, Andreas Veglis, George Kalliris, and Charalambos Dimoulas. 2013. "Journalism in a State of Flux: Journalists as Agents of Technology Innovation and Emerging News Practices." *International Communication Gazette* 75 (1): 76–98.
- Thurman, Neil. 2011. "Making 'the Daily Me': Technology, Economics and Habit in the Mainstream Assimilation of Personalized News." *Journalism: Theory, Practice & Criticism* 12 (4): 395–415.
- Thurman, Neil, Steve Schifferes, Richard Fletcher, Nic Newman, Stephen Hunt, and Aljosha Karim Schapals. 2016. "Giving Computers a Nose for News: Exploring the Limits of Story Detection and Verification." *Digital Journalism* 4 (7): 838–848.
- Ward, Stephen J. A. 2010. *Global Journalism Ethics*. Montreal: McGill-Queen's University Press.
- Weischenberg, Siegfried, Maja Malik, and Armin Scholl. 2006. *Die Souffleure der Mediengesellschaft: Report über die Journalisten in Deutschland* [The Prompters of the Media Society: Report on Journalists in Germany]. Konstanz: UVK.
- Weiss, Robert S. 1994. *Learning from Strangers*. New York: The Free Press.

- Willnat, Lars, David H. Weaver, and Jihyang Choi. 2013. "The Global Journalist in the Twenty-First Century: A Cross-National Study of Journalistic Competencies." *Journalism Practice* 7 (2): 163–183.
- Young, Mary Lynn, and Alfred Hermida. 2015. "From Mr. and Mrs. Outlier to Central Tendencies." *Digital Journalism* 3 (3): 381–397.
- Zion, Lawrie, and David Craig. 2014. *Ethics for Digital Journalists: Emerging Best Practices*. New York: Routledge.

Neil Thurman, (author to whom correspondence should be addressed), Department of Communication Studies and Media Research, LMU Munich, Germany and City, University of London, UK. E-mail: neil.thurman@ifkw.lmu.de. Twitter: [@neilthurman](https://twitter.com/neilthurman)

Konstantin Dörr, Institute of Mass Communication and Media Research, University of Zurich, Switzerland. E-mail: k.doerr@ipmz.uzh.ch. Twitter: [@kndoerr](https://twitter.com/kndoerr)

Jessica Kunert, Department of Communication Studies and Media Research, LMU Munich, Germany. E-mail: jessica.kunert@ifkw.lmu.de. Twitter: [@jessicakunert](https://twitter.com/jessicakunert)

Artikel IV

Normative Qualitätsansprüche an algorithmischen Journalismus

Zusammenfassung

Die fortschreitende Digitalisierung und Automatisierung im Journalismus setzen auch professionsethisches Handeln unter Druck. Dabei erzeugen die automatisierte Textproduktion und deren Verbreitung ethische Fragen auf individueller, organisatorischer und gesellschaftlicher Ebene. Bezogen auf normative Qualitätsansprüche diskutiert der Beitrag Fragen der journalistischen Verantwortung und der Transparenz. Es zeigt sich, dass die Technologie die Dimensionen der Verantwortung vor, während und nach der Nachrichtenproduktion maßgeblich beeinflusst. Neue Akteure spielen dabei ebenso eine zentrale Rolle wie die normativen Ansprüche im Umgang und bei der Kennzeichnung algorithmisch generierter Inhalte.

Dörr, K., Köberer, N., & Haim, M. 2017. Normative Qualitätsansprüche an algorithmischen Journalismus. In *Gesellschaft ohne Diskurs? Digitaler Wandel und Journalismus aus medienethischer Perspektive. Kommunikations- und Medienethik*, 5, herausgegeben von Stapf, I., Prinzing, M., & Filipović, A., 121–131. Baden-Baden: Nomos.

Copyright: Nomos Verlagsgesellschaft mbH & Co. KG

Genehmigung zum Druck wurde erteilt:

Dr. Sandra Frey

Nomos Verlagsgesellschaft mbH & Co. KG

Lektorat Soziologie, Medien- und Kommunikationswissenschaft

Normative Qualitätsansprüche an algorithmischen Journalismus

Konstantin Dörr, Nina Köberer, Mario Haim

Die fortschreitende Digitalisierung und Automatisierung im Journalismus stellt auch professionsethisches Handeln unter Druck. Dabei erzeugen die automatisierte Textproduktion und deren Verbreitung ethische Fragen auf individueller, organisatorischer und gesellschaftlicher Ebene. Bezogen auf normative Qualitätsansprüche diskutiert der Beitrag Fragen der journalistischen Verantwortung und der Transparenz. Es zeigt sich, dass die Technologie die Dimensionen der Verantwortung vor, während und nach der Nachrichtenproduktion maßgeblich beeinflusst. Neue Akteure spielen dabei ebenso eine zentrale Rolle wie die normativen Ansprüche im Umgang und bei der Kennzeichnung algorithmisch generierter Inhalte.

„Würdest du mir bitte sagen, wie ich von hier aus weitergehen soll?“
 „Das hängt zum großen Teil davon ab, wohin du möchtest“ sagte die Katze.
Lewis Carroll, Alice im Wunderland

1. Journalismus unter Druck: Algorithmen im Anmarsch

An den Eckpfeilern der professionellen Nachrichtenproduktion hat sich in den letzten Jahren wenig geändert: Die wirtschaftliche Lage vieler Medienorganisationen ist unsicher und wettbewerbsintensiv und treibt diverse Diskussionen voran. Die Frage, inwiefern journalistische Berichterstattung durch „User-Generated Content“ ergänzt werden kann und welche Folgen dies hat (vgl. dazu Köberer / Sehr in diesem Band), ist ebenso relevant wie die Diskussion über die zunehmende Automatisierung. Algorithmen klassifizieren, strukturieren und filtern – je nach Funktion der Anwendung – digitale Informationen und weisen ihnen auf Basis programmierter Regeln Relevanz zu (vgl. Rieder 2004). Für den Journalismus zeigt sich dieser Einfluss u. a. bei der Suche von Inhalten (Google Search), bei der Aggregation (Google News), der Empfehlung oder auch bei der algorithmischen Textproduktion (vgl. Latzer et al. 2014).

Diese Prozesse und die neuen Möglichkeiten beim Zugang und der Speicherung digitaler Daten (Big Data) verändern die Art und Weise, wie Journalismus die Gesellschaft beobachten und beschreiben kann (vgl. u. a.

Konstantin Dörr, Nina Köberer, Mario Haim

Napoli 2014; Neuberger / Nuernbergk 2015; Just / Latzer 2016). Die Entwicklung setzt damit auch professionsethisches Handeln unter Druck, da Handlungsorientierungen und -anweisungen für den Umgang mit und die Produktion von journalistischen Inhalten auf organisatorischer wie individueller Ebene kontinuierlich angepasst werden müssen (z. B. Ethik-Kodizes). Wie kann Journalismus seine Qualitätsansprüche in diesem Spannungsfeld behaupten? Und wem ist, vor dem Hintergrund neuer Akteure auf dem Nachrichtenmarkt, diese Verantwortung zuzuschreiben (vgl. u. a. Köberer 2014; Culver 2016)?

Dieser Beitrag diskutiert die ethischen Herausforderungen der automatisierten Textproduktion im Hinblick auf bestehende normative Qualitätsansprüche. Er verortet diese auf individueller, organisatorischer und gesellschaftlicher Ebene (vgl. Weischenberg / Malik / Scholl 2006). Die Form der automatisierten Inhalterstellung wird hier als *algorithmischer Journalismus* operationalisiert (vgl. u. a. Dörr 2015). Dabei wird die Medienethik, explizit die journalistische Ethik, als Bezugspunkt für die Diskussion herangezogen. Der Beitrag gliedert sich in drei Teile. Zunächst wird algorithmischer Journalismus innerhalb der Forschung verortet. Der anschließenden Diskussion der Kriterien normativer Qualität im Journalismus folgt die Auseinandersetzung mit der Frage nach *Verantwortung* und *Transparenz* als Qualitätskriterien im algorithmischen Journalismus. Das Fazit überführt die Ergebnisse in Vorschläge für einen künftigen Umgang mit dieser neuen Form des Journalismus.

2. *Algorithmischer Journalismus als Bereich des Computational Journalism*

In der Praxis hat sich das journalistische Handlungsfeld in vielfacher Weise ausdifferenziert. Der Einfluss von Algorithmen und Big Data verändert nicht nur das journalistische Aufgabenfeld, sondern auch die journalistischen Rollen vor, während und nach der Nachrichtenproduktion. Nach wie vor ist es Aufgabe des Journalismus auf *gesellschaftlicher Ebene*, Teilbereiche der Gesellschaft zu beobachten und Öffentlichkeit herzustellen (vgl. Weischenberg / Malik / Scholl 2006: 346). Darüber hinaus umfasst die definitorische Eingrenzung von Journalismus auch die Herstellung von Aussagen. Auf dieser *organisatorischen Ebene* werden die Inhalte in Redaktionen nach spezifischen Regeln und Routinen und damit professionell produziert. Hier schließt die dritte Ebene des Journalismus an, die der *pro-*

professionellen Akteure, welche über journalistische Arbeitsrollen und der Orientierung an professionsethischen Kriterien an der Produktion der Medienprodukte mitwirken (vgl. ebd.: 347). Auch algorithmischer Journalismus kann klassische Aufgaben und Funktionen des Journalismus erfüllen und eröffnet damit einen weiteren Spielraum für die Diskussion der ethischen Herausforderungen (vgl. Haim / Dörr 2016; Dörr / Hollnbuchner 2016; Ananny 2016). Doch auf welchen Ebenen werden sie sichtbar und was ist überhaupt algorithmischer Journalismus?

Die Funktionsweise von algorithmischem Journalismus lässt sich in Anlehnung an Dörr (2015) wie folgt beschreiben: Strukturierte, digitale Daten werden von Algorithmen selektiert, gewichtet und es wird ihnen je nach Programmierung und Beschaffenheit der Datensets Relevanz (Priorisierung und Kontextualisierung der Daten) zugewiesen. Die Daten werden also zunächst in eine semantische Struktur gebracht und dann in Textform in natürlicher und grammatikalisch korrekter Sprache über spezifische Kanäle auf einer reichweitenrelevanten Plattform durch einen professionellen journalistischen Akteur publiziert. Nach initialer Programmierung kann das System relativ autonom, effizient (schnell) und effektiv (fehlerfrei) journalistische Texte erstellen, publizieren und distribuieren.

Im wissenschaftlichen Diskurs werden die technischen, ökonomischen, rechtlichen und ethischen Herausforderungen der automatisierten Texterstellung auch unter den Begriffen „Robot Journalism“ (Clerwall 2014), „Automated Journalism“ (u. a. Carlson 2015; Graefe 2016) oder eben „Algorithmic Journalism“ (u. a. van Dalen 2012; Dörr 2015) diskutiert. Ein großer Teil der Forschung konzentriert sich dabei bislang auf die Wahrnehmung der Technologie durch Journalisten (u. a. Carlson 2014; van Dalen 2012) sowie durch Rezipierende (u. a. Clerwall 2014; Graefe et al. 2016). Grundsätzlich ist der algorithmische Journalismus dem „Computational Journalism“ zuzuordnen (vgl. Berger et al. 2015), worunter computergestützte Prozesse bei der Recherche, Datenanalyse, Visualisierung und Distribution, aber auch der Inhaltsproduktion gefasst werden (vgl. Hamilton / Turner 2009; Flew et al. 2012).

Nachrichtenmedien wie das Wirtschaftsmagazin *Forbes*, die *Los Angeles Times* (z. B. Quakebot für Erdbebennachrichten), *Bloomberg* oder die *Associated Press* binden diese automatisierten Inhalte bereits in ihr Angebot ein. Auch in Deutschland experimentieren Tageszeitungen wie die *Berliner Morgenpost* (mit dem Feinstaubmonitor) oder die *Stuttgarter Zeitung* mit solchen Software-Anwendungen. Dabei übernimmt algorithmischer Journalismus hauptsächlich im faktenbasierten Nachrichtenjourna-

Konstantin Dörr, Nina Köberer, Mario Haim

lismus auf Grundlage verfügbarer Daten journalistische Einzelleistungen. Die Darstellung von Meinung und die Einordnung von Ereignissen in gesellschaftliche Gesamtzusammenhänge kann diese Form des Journalismus nicht leisten. Es geht hier auch weniger um die Substitution des Journalisten und seines Leistungsspektrums. Vielmehr ermöglicht diese Technologie die massenhafte Textproduktion für spezielle Themenbereiche, um auch kleine Teilpublika mit Nachrichten zu versorgen (vgl. Dörr 2015). Diese sind aufgrund der eingeschränkten Datenverfügbarkeit aktuell zumeist auf Sport, Finanzen, Wetter und Verkehr beschränkt. Allmählich werden aber auch politische Inhalte abgedeckt (u. a. Haim / Graefe 2016). Aus medienethischer Perspektive sind dabei die Kriterien normativer Qualität (z. B. Transparenz) gleichwertig anzulegen, jedoch unabhängig von der thematischen Vielfalt.

3. Eine Frage von Verantwortung: Aspekte normativer Qualität

Vor dem Hintergrund einer zunehmenden Ausdifferenzierung des journalistischen Handlungsfeldes, der Ergänzung durch algorithmischen Journalismus und einer damit einhergehenden Veränderung journalistischer Aufgaben- und Tätigkeitsfelder, stellen sich auch Fragen neu, wem Verantwortung zugeschrieben wird und wer sie wie tragen soll. Dabei wird im medienethischen Diskurs auf einen Verantwortungsbegriff zurückgegriffen, bei dem verschiedene Ebenen berücksichtigt werden (vgl. Debatin 1998; Rath 2003a; Funiok 2007). Dazu zählen die Akteure, die einen direkten Umgang mit den Medien haben – *Medienmacher*, die *Besitzer und Betreiber von Massenmedien* sowie die *Mediennutzer*. Ergänzt werden diese durch die *freiwillige Selbstkontrolle der Medien*, die *medienkritische Öffentlichkeit* und den Bereich der *institutionellen Kontrolle und Gestaltung* (vgl. Debatin 1998: 121f.). Die Differenzierung des Mediensystems in diverse Handlungs- und Verantwortungsbereiche ermöglicht dabei die Verteilung und Zuschreibung von Verantwortung auf individuelle sowie auf sozioethische Ebene und bezieht sich damit auf einzelne Medienakteure sowie auf Akteure auf organisatorischer und gesellschaftlicher Ebene.

Bei Fragen nach Verantwortung und normativer Qualität im journalistischen Handlungsfeld stellt die journalistische Ethik einen wesentlichen Bezugspunkt dar. Im Kontext journalistischer Ethik wird versucht, Normen journalistischen Handelns zu begründen und Handlungsorientierung

Normative Qualitätsansprüche an algorithmischen Journalismus

für die journalistische Praxis zur Verfügung zu stellen. Die normative Beurteilung journalistischen Handelns bezieht sich in der journalistischen Ethik immer auch auf die Sicherung journalistischer Qualität. Als wesentliches Element journalistischer Ethik ist dementsprechend zu klären, was unter journalistischer Qualität zu verstehen ist und wie diese im Kontext der Produktion, Distribution und Rezeption journalistischer Produkte umgesetzt werden kann. Da es den einen verbindlichen Qualitätsmaßstab nicht gibt, geht es bei der Bestimmung medialer Qualität immer auch um die jeweils spezifisch zu thematisierenden und offenzulegenden Kriterien, Ziele und Inhalte des Begriffs (vgl. Karmasin 2005: 21). Neben dem Kriterium der Funktionalität äußert sich mediale Qualität auch über die Kriterien „Form und Inhalt bzw. Wirkung“ (ebd.). Bezugspunkt in ästhetischer Hinsicht ist die Form, also die äußere Erscheinung, wobei auch eine Ästhetik der Unterhaltung in diesem Rahmen ihre Berechtigung hat. Die inhaltliche Qualität bemisst sich an moralischen und ethischen Prinzipien, so dass „die ethisch/moralische Qualität medialer Interaktion an der Übereinstimmung mit den Normen medialer Ethik gemessen“ werden kann (ebd.: 25). Qualität im Journalismus beschreibt die Passung journalistischer Leistungen auf gesetzte Qualitätsansprüche. Ob eine journalistische Leistung als qualitativ hochwertig bezeichnet werden kann, hängt davon ab, welche Ansprüche an sie gestellt werden. Diese Ansprüche können unterschiedlich ausgestaltet (z. B. normativ, an Rezipienten orientiert, auf bestimmte Formate bezogen) und an unterschiedliche Instanzen gerichtet (z. B. Beiträge, Personen, Redaktionen) sein (vgl. u. a. Arnold 2008; Hohlfeld 2010; Jungnickel 2011; Neuberger 2014; fög 2015).

Mit Blick auf das journalistische Handlungsfeld und die Frage, welche journalistischen Qualitätsmaßstäbe Geltung beanspruchen, muss daher zunächst zwischen den Aufgaben und Funktionen journalistischer Berichterstattung differenziert werden. Im Informationsbereich zählen ethische Kriterien wie die *Achtung vor der Wahrheit*, die *Wahrung der Menschenwürde* und die *wahrhaftige Unterrichtung der Öffentlichkeit* zu den Grundnormen des Journalismus (vgl. Wunden 2003: 175ff.). Nur die wahrhaftige, sachgerechte und verständliche Vermittlung von Informationen ermöglicht eine tatsächliche Orientierung in diversen Sachverhalten und ist eine wichtige Voraussetzung für die freie Meinungsbildung, die Partizipation an Öffentlichkeit und damit eine Grundbedingung für die Umsetzung demokratischer Prinzipien. Als Prinzipien journalistischer Qualitätssicherung lassen sich mit Bezug auf die Transparenz auch die von Pöttker (1999) formulierten Trennungsgrundsätze nennen. Ihm zufolge stellen die *Trennung*

Konstantin Dörr, Nina Köberer, Mario Haim

von *Information und Werbung*, die *Trennung von Information und Fiktion* sowie die *Trennung von Information und Meinung* Grundvoraussetzungen für die Glaubwürdigkeit der Medien und die Qualität journalistischer Berichterstattung dar. Sofern die Frage journalistischer Qualität sich auf Informationsprodukte bezieht, ist die Forderung nach Qualitätsmerkmalen wie *Wahrhaftigkeit* und Transparenz aus normativer Perspektive weiterhin zu fordern (vgl. Rath 2006; Köberer 2014; Ward 2014). Diese typischen Normen zur Verhaltensorientierung finden sich auch in den Publizistischen Grundsätzen (z. B. im Pressekodex des Deutschen Presserats oder auch im Ehrenkodex des österreichischen Presserats).

Die Diskussion ethischer Herausforderungen im Journalismus ist immer auch geprägt durch individuelle, ökonomische und strukturelle Veränderungen, die sich durch neue Techniken entwickeln. Das Zusammenspiel von Technik und Journalismus ist hierbei als ko-evolutionärer Prozess zu verstehen, der den Fokus auf den Prozesscharakter medialen Wandels legt (vgl. Latzer 2013). Da algorithmischer Journalismus bereits journalistische Leistungen übernehmen kann bzw. diese in der Praxis auch schon übernimmt und damit in das Angebot professioneller Akteure integriert ist, müssen hier ebenfalls journalistische Qualitätsansprüche angelegt werden (Haim / Dörr 2016). Herausforderungen zeigen sich dabei auf der Ebene der Datenauswahl, -nutzung, -suche und -herkunft (Zion / Craig 2014), der Objektivität (Gillespie 2014), der Transparenz (McBride / Rosenstiel 2013), und auf der Ebene der Implementierung von Werten im Programmcode (Kraemer / van Overveld / Peterson 2011; Young / Hermida 2015).

Nachfolgend soll zunächst die übergeordnete Frage nach der *Verantwortung* von Akteuren im Produktionsprozess geklärt werden, bevor exemplarisch *Transparenz* als Wert medialer Qualität im Kontext algorithmischer Inhaltsproduktion diskutiert wird.

3.1 Neue Akteure und ihre Verantwortung im algorithmischen Journalismus

Im medienethischen Diskurs sind sowohl Einzelpersonen (z. B. Medienmacher, Mediennutzer) als auch Unternehmen und Medienorganisationen (z. B. Zeitungsverlage) bzw. Redaktionen, die journalistische Inhalte anbieten, als Akteure zu verorten. Durch die stetige Automatisierung der Produktionsprozesse im Journalismus ist mit Blick auf die Frage der Zuschreibung und Übernahme von Verantwortung aus ethischer Perspektive

Normative Qualitätsansprüche an algorithmischen Journalismus

zunächst zu klären, ob bzw. inwiefern algorithmischer Journalismus bzw. Algorithmen als Träger von Verantwortung verstanden werden können. Grundsätzlich waren und sind ethische Fragestellungen immer subjektgebunden. Nur der Mensch kann rational und vernunftbegabt handeln (z. B. Kant KpV 1788). Mit automatisiert handelnden Algorithmen, die immer unabhängiger in ihren Entscheidungs- und Bewertungsmöglichkeiten werden, ist nunmehr neben nicht journalistisch sozialisierten Akteuren wie Software-Anbietern oder Programmierern auch verstärkt die Rolle der Technik als Akteur in die ethische Diskussion einzubeziehen. Dabei stellen sich grundsätzliche Fragen nach einer möglichen Handlungsträgerschaft von Algorithmen und der Option, moralisch handeln zu können (vgl. Kroes / Verbeek 2014). Mit Rückgriff auf techniksoziologische Ansätze zeigt sich dabei für den algorithmischen Journalismus, dass eine vollautomatisierte Textproduktion und eine Kontextualisierung dieser Inhalte ohne menschliche Beteiligung noch nicht möglich ist (z. B. bei der Themenauswahl, Interpretation und Einordnung in gesellschaftliche Zusammenhänge) (vgl. Dörr 2015). Doch erst diese Anlehnung an andere theoretische Konzepte macht aktuelle technische Fragen einer breiteren normativen Analyse zugänglich. Denn die steigende Komplexität der Interaktion zwischen Mensch und Algorithmus erhöht auch die an *Medienmacher* und *Medienbesitzer* zu stellenden Anforderungen. Sie müssen auch in diesem Wechselspiel die zentralen Leistungen des Journalismus garantieren.

Das Zusammenwirken von Algorithmen und Journalisten im algorithmischen Journalismus ordnet ferner die Verantwortungsebenen vor, während und nach der Produktion neu (vgl. Dörr / Hollnbuchner 2016). Der Journalist ist dabei nicht mehr die alleinige moralische Instanz bei der Auswahl der Daten und der Verarbeitung. Die Technologie ist zu komplex und das Technikverständnis vieler Journalisten zu gering, um die komplexen Selektionsentscheidungen der Algorithmen verstehen zu können. Konkurrenz in seiner Funktion als Nachrichtenproduzent bekommt der Journalist durch Software-Anbieter, die den Zugang zu der Technologie ermöglichen und das automatisierte journalistische Produkt im Auftrag einer Medienorganisation programmieren und anbieten. Sie verarbeiten die Daten, die von Medienorganisationen gestellt werden, und programmieren die spezifische Anwendung je nach Zweck. So entsteht etwa eine Anwendung, die zwar autonom Zusammenfassungen der EM-Spiele der deutschen Nationalmannschaft schreiben kann, jedoch nicht ohne weiteres auf die WM oder gar die Bundesliga zu übertragen ist. Da eine Kontrolle

Konstantin Dörr, Nina Köberer, Mario Haim

der Texte oft aber erst nach Veröffentlichung stattfinden kann, stellt sich die Frage, ob diese inhaltliche Verantwortung vor dem Hintergrund der kommerziellen Verwert- und Skalierbarkeit erhalten bleibt oder neuen Akteuren zugeschrieben werden muss (vgl. Dörr / Hollnbuchner 2016). Dabei lassen sich Rollenverschiebungen und -erweiterungen nicht nur zwischen Journalist und Software-Anbieter „diskutieren“, sondern es müssen künftig auch die Einflüsse von Algorithmen, Programmierern und Publikum auf den Produktionsprozess und auf die Leistungserfüllung einbezogen und genauer untersucht werden (vgl. Lewis / Westlund 2015): Wie verändern sich normative Ansprüche, wenn nicht-journalistisch sozialisierte Akteure wie Programmierer, Software- oder Daten-Anbieter eine zentrale Rolle bei der Nachrichtenproduktion spielen? Ethische Richtlinien und Kodizes für den Umgang von automatisierten Inhalten auf organisatorischer Ebene sind derzeit noch Mangelware. Stattdessen wird lediglich, wenngleich konsequent auf Transparenz gedrängt (u. a. Diakopoulos 2015; Dörr / Hollnbuchner 2016), die sowohl algorithmischer Journalismus im Speziellen als auch Computational Journalism als Ganzes häufig vermissen lässt (Pasquale 2015).

3.2 Transparenz und algorithmischer Journalismus

Transparenz gilt als zentrales Kriterium normativer Fragestellungen (u. a. Jungnickel 2011), auch und vor allem im algorithmischen Journalismus (vgl. Dörr / Hollnbuchner 2016). Aus medienethischer Perspektive sollten automatisierte Inhalte ebenso als solche gekennzeichnet werden, wie im Sinne des Trennungsgrundsatzes beispielsweise alle Werbeinhalte, die wie redaktionelle Beiträge gestaltet und nicht als solche erkennbar sind, als Werbung ausgewiesen werden müssen (vgl. Köberer 2014; Haim / Dörr 2016). Die Erkennbarkeit der Autorenschaft eines Beitrags ist dabei nicht nur wichtig für den Anspruch an Transparenz als Kriterium journalistischer Qualität, sondern lässt auch erste Rückschlüsse über die Interessen eines Anbieters erkennen. Obschon einige Medienorganisationen (z. B. die Agentur AP) ihre Inhalte korrekt kennzeichnen, bleibt die Frage, was dem Anspruch von Transparenz genügt: die Offenlegung der Automatisierung, die Offenlegung der Datenquellen oder gar die Offenlegung des Programmcodes? Zumindest bei Letzterem dürften wohl die Software-Anbieter um ihr Betriebsgeheimnis fürchten. Denn der verstärkte Wettbewerb um die Märkte hat zu veränderten Prioritäten und Funktionen der Medien-

Normative Qualitätsansprüche an algorithmischen Journalismus

angebote und damit auch zu einer Veränderung der journalistischen Angebote geführt. Aus ökonomischer Sicht liegt es nicht im Interesse eines Medienunternehmens, „ein bestimmtes Medienprodukt zu produzieren, sondern einen Gewinn zu erwirtschaften“ (Rath 2003b: 133). Das Produkt dient als Mittel zum Zweck und oftmals ist die Qualität des Produkts nur insofern relevant, als dass das Produkt eine bestimmte Qualitätsvorstellung des Konsumenten (in diesem Fall: Rezipienten) zu erfüllen hat, um überhaupt konsumiert zu werden. „Qualität ist eine zentrale Kategorie des Vermarktungserfolgs“ von Medienprodukten und wird, ökonomisch betrachtet, an Merkmalen wie der Kundenorientierung und der Zweckmäßigkeit festgemacht (ebd.: 133). Diese Kriterien reichen jedoch nicht aus, um einen Begriff von medialer Qualität zu bestimmen, der über eine Mittel-Zweck-Beziehung hinausgeht und normative Orientierung bietet. Nicht alles, was marktfähig gemacht werden kann, ist aus normativer Perspektive auch markttauglich. Insbesondere Fragen der Transparenz konterkarieren diese ökonomische Sichtweise. Die ethischen Herausforderungen zeigen sich beim Datenzugang und der Datenauswahl: Wie sachgerecht und vollständig sind die Datensets, die zur Texterstellung genutzt werden? Woher kommen die Daten, wer hat sie gesammelt? Können sie das abbilden, was berichtet werden soll, oder ist der finale Text verzerrt, weil beispielsweise wichtige Informationen im Datensatz fehlen? Hier stellen sich Fragen zum Quellenschutz, die auch hinsichtlich dieser Kriterien abgewogen werden müssen.

4. Ausblick: Implikationen für den Umgang mit algorithmischen Journalismus

Die Wirkungsqualität medialer Angebote normativ abzuschätzen, also einer rationalen Reflexion zu unterziehen und nach wissenschaftlich plausiblen, normativen Kriterien zu bewerten, ist von der Medienethik zu leisten. Normative Qualitätsansprüche an den Journalismus aus medienethischer Perspektive sind auch auf neue Phänomene wie den algorithmischen Journalismus zu übertragen und inhaltlich weiterzudenken.

Algorithmen sind zwar qua definitionem prädestiniert dafür, repetitiv und nachvollziehbar, fehlerfrei und ohne Verzerrungen zu arbeiten, gesetzten Vorgaben Folge zu leisten und dabei auf die formalen Bedürfnisse der Rezipienten einzugehen (vgl. Haim / Dörr 2016; Graefe et al. 2016), dennoch ergeben sich insbesondere auf der Ebene des Datenzugangs/-auswahl

Konstantin Dörr, Nina Köberer, Mario Haim

und der Ebene des Programmcodes (Komplexität) neue normative Herausforderungen. Mit der Automatisierung der Nachrichtenproduktion zeigt sich insbesondere auch eine Verantwortungsdiffusion im journalistischen Gestaltungsprozess. In diesen Prozess sind neben Journalisten, die etwa für die Konfiguration der Algorithmen verantwortlich sind, Programmierer externer Dienstleister für deren Entwicklung sowie Datenanbieter für die fehlerfreie und vollständige Informationsgrundlage eingebunden (vgl. Dörr 2015). Die einzelnen Akteure verfolgen dabei mitunter konterkarierende Geschäftsinteressen, sodass Verantwortlichkeiten delegiert und Arbeitsschritte hinter verschlossenen Türen durchgeführt werden. In letzter Instanz müssen die automatisierten Texte kommerzielle Ziele von Medienorganisationen erfüllen können, z. B. durch den Verkauf von Werbung, Ressourceneinsparungen oder der Gewinnung neuer Abonnenten. Nur wenn auch diese Kriterien erfüllt werden, haben Medienorganisationen ein langfristiges Interesse an der Entwicklung und Verbreitung der Technologie im Journalismus.

Eine differenzierte Bewertung der unterschiedlichen Akteure ist notwendig und mit Blick auf den rasanten Fortschritt im Bereich der Informationstechnologie einem entsprechend schnellen Wandel ausgesetzt. Es schließen sich insbesondere medienregulatorische Fragestellungen an (vgl. Saurwein / Just / Latzer 2015). Auch die Rolle des Publikums ist zu berücksichtigen, da es (indirekt) über die Kriterien normativer Qualität mitentscheidet, indem Texte auch trotz mangelnder Kennzeichnung geklickt und gelesen werden.

Es bleibt die gleichermaßen ernüchternde wie perspektivische Erkenntnis, dass Qualitätsansprüche auch für algorithmischen Journalismus zunächst der Sache gerecht werden müssen. Aber ähnlich wie Alice im Wunderland steht auch der Journalismus vor der Frage wie es von hier aus weitergehen soll. Denn erst das Zusammenspiel professioneller und nicht-professioneller Akteure auf organisatorischer und individueller Ebene, die Berücksichtigung der Rolle von Algorithmen sowie die Integration des Publikums, kann die traditionellen Kriterien normativer Qualität bezogen auf algorithmischen Journalismus festigen oder auch neu verhandeln.

Literatur

- Ananny, Mike (2016): *Toward an Ethics of Algorithms: Convening, Observation, Probability, and Timeliness*. In: *Science, Technology, & Human Values* 41 (1), S. 93–117.
- Arnold, Klaus (2008): *Qualität im Journalismus. Ein integratives Konzept*. In: *Publizistik* 53 (4), S. 488–508.
- Berger, Matthias / Haim, Mario / Graefe, Andreas / Brosius, Hans-Bernd / Hess, Thomas (2015): *Aktuelles Stichwort: Computational Journalism*. In: *Medienwirtschaft* 12 (1), S. 20–23.
- Brosda, Carsten (2010): *Journalismus*. In: Brosda, Carsten / Schicha, Christian: *Handbuch Medienethik*. Wiesbaden: VS Verlag, S. 257–277.
- Carlson, Matt (2015): *The Robotic Reporter*. In: *Digital Journalism* 3, S. 416–431.
- Clerwall, Christer (2014): *Enter the Robot Journalist. Users' perceptions of automated content*. In: *Journalism Practice* 8 (5), S. 519–531.
- Culver, Kathleen Bartzen (2016): *Disengaged Ethics. Code development and journalism's relationship with 'the public'*. In: *Journalism Practice*, S. 1–16.
- Debatin, Bernhard (1998): *Verantwortung im Medienhandeln. Medienethische und handlungstheoretische Überlegungen zum Verhältnis von Freiheit und Verantwortung in der Massenkommunikation*. In: Wunden, Wolfgang (Hg.): *Freiheit und Medien*. Frankfurt a. M.: Abt. Verlag (Beiträge zur Medienethik. Band 4.), S. 113–130.
- Diakopoulos, Nicholas (2015): *Algorithmic Accountability. Journalistic investigation of computational power structures*. In: *Digital Journalism* 3 (3), S. 398–415.
- Dörr, Konstantin (2015): *Mapping the Field of Algorithmic Journalism*. In: *Digital Journalism*, S. 1–24.
- Dörr, Konstantin / Hollnbuchner, Katharina (2016): *Ethical Challenges of Algorithmic Journalism*. In: *Digital Journalism*, S. 1–16.
- Flew, Terry / Spurgeon, Christina / Daniel, Anna / Swift, Adam (2012): *The promise of computational Journalism*. In: *Journalism Practice* 6 (2), S. 157–171.
- Fög (Forschungsinstitut Öffentlichkeit und Gesellschaft – Universität Zürich) (Hg.) (2015): *Jahrbuch Qualität der Medien 2015 Schweiz*. Basel: Schwabe.
- Funiok, Rüdiger (2007): *Medienethik: Verantwortung in der Mediengesellschaft*. Stuttgart: Kohlhammer.
- Gillespie, Tarleton (2014): *The relevance of algorithms*. In: Gillespie, Tarleton / Boczkowski, Pablo J. / Foot, Kirsten A. (Hg.): *Media technologies. Essays on communication, materiality, and society*. Cambridge, MA: MIT Press, S. 167–193.
- Graefe, Andreas (2016): *Guide to Automated Journalism*. In: *Tow Center for Digital Journalism*.
- Graefe, Andreas / Haim, Mario / Haarmann, Bastian / Brosius, Hans-Bernd (2016): *Readers' perception of computer-generated news: Credibility, expertise, and readability*. In: *Journalism*, S. 1–16.

Konstantin Dörr; Nina Köberer; Mario Haim

- Haim, Mario / Dörr, Konstantin (2016): *Schema F. Zur Institutionalisierung und normativen Qualität von algorithmischem Journalismus*. Vortrag auf der Jahrestagung der DGPK, Leipzig.
- Haim, Mario / Graefe, Andreas (2016): *Automated news: Better than expected?* Vortrag auf der Jahrestagung der ICA, Fukuoka/Japan.
- Hamilton, James T. / Turner, Fred (2009): *Accountability Through Algorithm. Developing the Field of Computational Journalism. A report from Developing the Field of Computational Journalism*. Online verfügbar unter: <http://web.stanford.edu/~fturner/Hamilton%20Turner%20Acc%20by%20Alg%20Final.pdf> (Abfrage am: 30. April 2016).
- Hohlfeld, Ralf (2010): *Publizistische Qualität in neuen Öffentlichkeiten. Crossmedia als Herausforderung für die Verbindung einer Theorie publizistischer Qualität mit einer Theorie publizistischer Innovation*. In: Hohlfeld, Ralf / Müller, Phillip / Richter, Annekathrin / Zacher, Franziska (Hg.): *Crossmedia – Wer bleibt auf der Strecke?* Bd. 1. Berlin: LIT., S. 20–36.
- Jungnickel, Katrin (2011): *Nachrichtenqualität aus Nutzersicht. Ein Vergleich zwischen Leserurteilen und wissenschaftlich-normativen Qualitätsansprüchen*. In: *Medien & Kommunikationswissenschaft* 59 (3), S. 360–378.
- Just, Natascha / Latzer, Michael (2016): *Governance by algorithms: reality construction by algorithmic selection on the Internet*. In: *Media, Culture & Society*, S. 1–21.
- Kant, Immanuel (1788): *Kritik der praktischen Vernunft*. Riga: Hartknoch.
- Karmasin, Matthias (2005): *Journalismus: Beruf ohne Moral? Von der Berufung zur Profession*. Wien: Facultas.
- Köberer, Nina (2014): *Advertorials in Jugendprintmedien. Ein medienethischer Zugang*. Wiesbaden: Springer VS.
- Kraemer, Felicitas / van Overveld, Klees / Peterson, Martin (2011): *Is there an ethics of algorithms?* In: *Ethics & Information Technology* 13 (3), S. 251–260.
- Kroes, Peter / Verbeek, Peter-Paul (2014): *The Moral Status of Technical Artefacts*. Wiesbaden: Springer.
- Latzer, Michael (2013): *Medienwandel durch Innovation, Ko-Evolution und Komplexität*. In: *Medien & Kommunikationswissenschaft* 61 (2), S. 235–252.
- Latzer, Michael / Hollnbuchner, Katharina / Just, Natascha / Saurwein, Florian (2014): *The economics of algorithmic selection on the Internet*. Online verfügbar unter: www.mediachange.ch/media/pdf/publications/Economics_of_algorithmic_selection_WP.pdf (Abfrage am: 20. April 2016).
- Lewis, Seth C. / Westlund, Oscar (2015): *Big Data and Journalism. Epistemology, Expertise, Economics, and Ethics*. In: *Digital Journalism* 3 (3), S. 447–466.
- McBride, Kelly / Rosenstiel, Tom (2013): *New Guiding Principles for a New Era of Journalism*. In: McBride, Kelly / Rosenstiel, Tom: *The New Ethics of Journalism*. Thousand Oaks, CA: CQ Press, S. 1–6.
- Napoli, Philip M. (2014): *Automated Media: An Institutional Theory Perspective on Algorithmic Media Production and Consumption*. In: *Communication Theory* 24 (3), S. 340–360.

Normative Qualitätsansprüche an algorithmischen Journalismus

- Neuberger, Christoph (2014): *Die Identität und Qualität des Journalismus im Internet aus der Sicht des Publikums*. In: Loosen, Wiebke / Dohle, Marco (Hg.): Journalismus und (sein) Publikum. Wiesbaden: VS Verlag, S. 229–251.
- Neuberger, Christoph / Nuernbergk, Christian (2015): *Verdatete Selbstbeschreibung der Gesellschaft – Über den Umgang des Journalismus mit Big Data und Algorithmen*. In: Süssenguth, Florian (Hg.): Die Gesellschaft der Daten – Über die digitale Transformation der sozialen Ordnung. Bielefeld: transcript, S. 199–224.
- Pasquale, Frank (2015): *The black box society: The secret algorithms that control money and information*. Cambridge, MA: Harvard University Press.
- Pöttker, Horst (1999): *Berufsethik für Journalisten? Professionelle Trennungsgrundsätze auf dem Prüfstand*. In: Holderegger, Adrian (Hg.): Kommunikations- und Medienethik. Interdisziplinäre Perspektiven. Freiburg: Universitätsverlag, S. 299–327.
- Rath, Matthias (2003a): *Medien in Zeiten der Globalisierung – Selbstregulierung zwischen Freiheit und Verantwortung*. In: Medienjournal 27 (1), S. 41–50.
- Rath, Matthias (2003b): *Die medienphilosophische Perspektive: Medien, Wirtschaft, Sinn*. In: Altmeyden, Klaus-Dieter / Karmasin, Matthias (Hg.): Medien und Ökonomie. Band 1/2. Grundlagen der Medienökonomie. Wiesbaden: VS Verlag, S. 125–139.
- Rath, Matthias (2006): *Wahrhaftigkeit des Journalismus: Moralanspruch oder Marktfaktor?* In: Niesyto, Horst / Rath, Matthias / Sowa, Hubert (Hg.): Medienkritik heute. Grundlagen, Beispiele und Praxisfelder. München: Kopaed, S. 117–128.
- Rieder, Bernhard (2004): *Algorithmische Mediatoren: Software-Agenten als Gegenstand der Medienwissenschaften*. In: Medien-Journal 28 (1), S. 36–46.
- Saurwein, Florian / Just, Natascha / Latzer, Michael (2015): *Governance of algorithms: options and limitations*. In: info 17 (6), S. 35–49.
- Van Dalen, Arjen (2012): *The algorithms behind the headlines. How machine-written news redefines the core skills of human journalists*. In: Journalism Practice 6 (5-6), S. 648–658.
- Ward, Stephen J. (2014): *The Magical Concept of Transparency*. In: Zion, Lawrie / Craig, David: Ethics for Digital Journalists: Emerging Best Practices. New York: Routledge, S. 45–58.
- Weischenberg, Siegfried / Malik, Maja / Scholl, Armin (2006): *Journalismus in Deutschland 2005. Zentrale Befunde der aktuellen Repräsentativbefragung deutscher Journalisten*. In: Media Perspektiven 7, S. 346–361.
- Wunden, Wolfgang (2003): *Die „Publizistischen Grundsätze“ des Deutschen Presse-rats aus medienethischer Sicht*. In: Debatin, Bernhard / Funiok, Rüdiger (2003) (Hg.): Kommunikations- und Medienethik. Konstanz: UVK, S. 169–182.
- Young, Mary L. / Hermida, Alfred (2015): *From Mr. and Mrs. Outlier To Central Tendencies*. In: Digital Journalism 3 (3), S. 381–397.
- Zion, Lawrie / Craig, David (2014): *Ethics for Digital Journalists. Emerging Best Practices*. New York: Routledge.

Zusammenfassung

With the institutionalization of algorithms as content creators, professional journalism is facing transformation and novel ethical challenges. This article focuses on the concept of Algorithmic Journalism on the basis of natural language generation and provides a framework to identify and discuss ethical issues. The analysis builds on the moral theories of deontology, utilitarianism, virtue ethics, and contractualism, and remaps the ethical discussion for Algorithmic Journalism at the intersection of digital media ethics and cyber ethics. In order to capture the whole range of potential shifts and challenges in journalism ethics, the article combines the ethical multi-layer system of responsibility by Pürer with the classification of journalism by Weischenberg, Malik, and Scholl on an organizational, professional/individual, and social/audience sphere. This analytical framework is then complemented with attributes derived from the technical potential of Algorithmic Journalism. As a result, the analysis uncovers new ethical challenges and shifts of responsibility in news production for journalism practice and journalism research at the levels of objectivity, authority, transparency, and at the level of implicit or explicit values.

Dörr, K., & Hollnbuchner, K. 2016. Ethical Challenges of Algorithmic Journalism. *Digital Journalism* [online first]. doi: [10.1080/21670811.2016.1167612](https://doi.org/10.1080/21670811.2016.1167612)

Copyright: <http://www.tandfonline.com>

The article is published as the version of record in Digital Journalism:
<http://www.tandfonline.com/doi/full/10.1080/21670811.2016.1167612>

ETHICAL CHALLENGES OF ALGORITHMIC JOURNALISM

Konstantin Nicholas Dörr and Katharina Hollnbuchner

With the institutionalization of algorithms as content creators, professional journalism is facing transformation and novel ethical challenges. This article focuses on the concept of Algorithmic Journalism on the basis of natural language generation and provides a framework to identify and discuss ethical issues. The analysis builds on the moral theories of deontology, utilitarianism, virtue ethics, and contractualism, and remaps the ethical discussion for Algorithmic Journalism at the intersection of digital media ethics and cyber ethics. In order to capture the whole range of potential shifts and challenges in journalism ethics, the article combines the ethical multi-layer system of responsibility by Pürer with the classification of journalism by Weischenberg, Malik, and Scholl on an organizational, professional/individual, and social/audience sphere. This analytical framework is then complemented with attributes derived from the technical potential of Algorithmic Journalism. As a result, the analysis uncovers new ethical challenges and shifts of responsibility in news production for journalism practice and journalism research at the levels of objectivity, authority, transparency, and at the level of implicit or explicit values.

KEYWORDS automated journalism; cyber ethics; digital ethics; ethics of algorithms; journalism ethics; new media ethics; robot journalism; value-sensitive design

Introduction: Changing Journalism

Journalism and its production routines and conditions have always been shaped and influenced by technology (Parry 2011; Pavlik 2000). With algorithms entering the stage of professional news production, distribution, and consumption, editorial structures and journalistic routines are changing significantly (Broussard 2014; van Dalen 2012; Napoli 2014; Pavlik 2013). Along with these technological and social developments, values and responsibilities assigned to journalism and journalists are changing accordingly (Brosda 2010; Culver 2016; Meier 2002, 117f.). Due to the progress in natural language generation (NLG)—an area of research in computational linguistics—large numbers of texts can be produced automatically from digital structured data (Reiter and Dale 2000). For example, the Associated Press uses this technology to automate their corporate earning stories with Wordsmith, a software tool developed by Automated Insights. This phenomenon of automated algorithmic content creation and its application and use in journalism is here conceptualized as Algorithmic Journalism (Dörr 2015). Terms like robot journalism, automated journalism, or machine-written journalism are often used similarly (Anderson 2012; Carlson 2014).

With algorithms entering professional news production, ethical challenges arise on various levels. They are detected on the level of data search and origin (Bradshaw 2014), algorithmic authority (Carlson 2014), algorithmic objectivity (Gillespie 2014), algorithmic transparency (McBride and Rosenstiel 2014a), data usage and abuse (Zion and Craig 2014), and on the level of values and reasoning embedded in code (Kraemer, van Overveld, and Peterson 2011; Young and Hermida 2014). Diakopoulos (2015) also discusses ethical challenges of algorithmic accountability with a focus on reporting.

There have been discussions about ethics of cyber technological systems in general (e.g. Friedman, Kahn, and Borning 2008; Kernaghan 2014; Spinello 2011; Tavani 2011), as well as in specific fields of application, e.g. high-frequency trading (Davis, Kumiega, and Vliet 2012) or autonomous vehicles (Lin 2015). These cyber technologies can be distinguished from other (precedent) computer technologies as they are uniquely fast, uniquely complex, uniquely coded, uniquely universal (global), and uniquely ubiquitous (see Ess 2009; Tavani 2011, 4). These features also apply to Algorithmic Journalism (for the technological potential, see Dörr 2015).

Still, there has not yet been a consistent remapping of media ethics and journalism ethics focusing on this phenomenon (for an overview, see Prinzing et al. 2015). Therefore, the first research question asks:

RQ1: Where can the ethical challenges of Algorithmic Journalism be theoretically located?

This article then provides a framework to identify and discuss the ethical issues of algorithmic content generation in journalism.

To outline and structure the whole range of shifts and challenges in journalism ethics, the framework is theoretically based on Weischenberg, Malik, and Scholl's (2006) classification of journalism as a social, organizational, and professional sphere. This structuring is then combined with the ethical multi-layer system of responsibility by Pürer (1992) to journalism ethics. To identify the relevant levels of ethical challenges before, during, and after news production, this analytical framework is also complemented by attributes derived from the technical potential of Algorithmic Journalism.

RQ2: Which ethical challenges of Algorithmic Journalism can be identified along the organizational, professional, and social sphere?

Ethical Localization of Algorithmic Journalism

To answer RQ1, the following section illustrates the theoretical localization of Algorithmic Journalism within ethical concepts.

Generally speaking, ethics is a sub-area of philosophy dealing with the morality of human action (Prinzing et al. 2015). Morality as "a system of rules for guiding human conduct, and principles for evaluating those rules" (Tavani 2011, 36) and ethical behavior should lead to the common wellbeing of all. Therefore, four moral theories help to analyze and cluster argumentation in ethical disputes. First, there are consequence-based ethical theories like utilitarianism (Bentham [1825] 2015; Mill 2006) asking what leads to the greatest amount of good for the greatest number of people affected by an action. Second, duty-based approaches of deontology going back to Kant ([1797]

2003) and Ross ([1930] 2003) are arguing that morality is based on the duties and obligations of human interaction. Third, contract-based theories relying on Hobbes' *Leviathan* (Darwall 2003) are speaking of establishing moral systems through societal contracts. Finally, approaches of virtue ethics (e.g. Aristotle 2003) focus on the character of a person asking, "What kind of person should I be?" These moral theories help to capture ethical issues in journalism ethics in general and when dealing with Algorithmic Journalism in particular. A similar approach was suggested by Ananny (2015) by combining deontology, utilitarianism, and virtue ethics while describing the application of these theories as being restricted to certain areas. In his understanding, deontology is solely found in policies and regulations, utilitarianism in technology development, and virtue ethics only on an individual level. But when discussing journalism ethics and the influence of automated content production, we argue for a broader approach. As Ananny's (2015) analytical model for algorithmic applications focuses on ethical questions of when, how, and for whom, this analysis also takes questions of source (by whom?) and motivation (for which reasons?) into account.

As ethical challenges have always been part of media research (Brosda 2010; Hömberg and Klenk 2014; Ward 2006), traditional media ethics are a form of applied ethics focusing on the professional mass media system with its daily routines and professional practices (Horner 2013, 5). Thereby media ethics do provide a proactive examination of possible conflicts along the technological and social development in journalism.

New developments like networked technologies, new intermediaries, the deprofessionalization in journalism, or new working routines established in the light of the internet and technological change are usually tagged with the terms "new media ethics" (Debatin 2010) or "digital media ethics" (Ess 2009). The term "new media ethics" remains unsatisfactory: first, the term "new" is time-dependent; e.g. in the 1950s, television was new, but that does not apply today. Even the internet as a primary object of new media ethics is not "new" as it is a widely diffused and applied technology (Debatin 2010, 318). Second, "media" still mostly refers to mass media although new technologies do not need to have the same features. So far, the term "digital media ethics" coined by Ess (2009) seems more promising for the ethical discussion of Algorithmic Journalism as it focuses on technological core features and does not rely on a description of a technology's innovative status.

Following this general classification, media ethics in social sciences have been influenced recently by applied fields of cyber ethics in computer science (Maner 1996; Spinello 2011). This term refers to the study of moral, legal, and social issues involving cyber technologies, such as privacy, property, free speech, etc., which can be understood according to Tavani (2011, 11) as "specific expression of core (traditional) moral notions, such as autonomy, fairness, justice, responsibility, and respect for persons."

Within cyber ethics, other applied fields of ethics have to be considered for the ethical discourse on algorithms in journalism and especially for Algorithmic Journalism. These are, in particular, "computer ethics" (Johnson 2009; Pierce and Henry 1996) which refer to moral issues concerning computing, as well as "information ethics" (Capurro 2007) which refer to moral issues concerning the flow and the processing of information. Furthermore, "internet ethics" (Langford 2000) dealing with ethical issues concerning the internet and focusing on connectivity and the technological fundamentals, or more recently, "machine ethics" or "robo ethics" (Shulman et al. 2009; Borenstein and

Pearson 2012) referring to securing the ethical behavior of artificial intelligence systems and agents have to be included. This structuring is closely linked to the discussion of algorithms as relatively autonomous actors with delegated (moral) agency (Just and Latzer, *forthcoming*).

The analysis of Algorithmic Journalism’s ethical implications is therefore situated in the overlapping fields of digital media ethics—as a part of journalism ethics—and cyber ethics. Figure 1 illustrates the localization of Algorithmic Journalism in the ethical field and highlights RQ1.

Journalism has been one of the first and best elaborated areas in media ethics

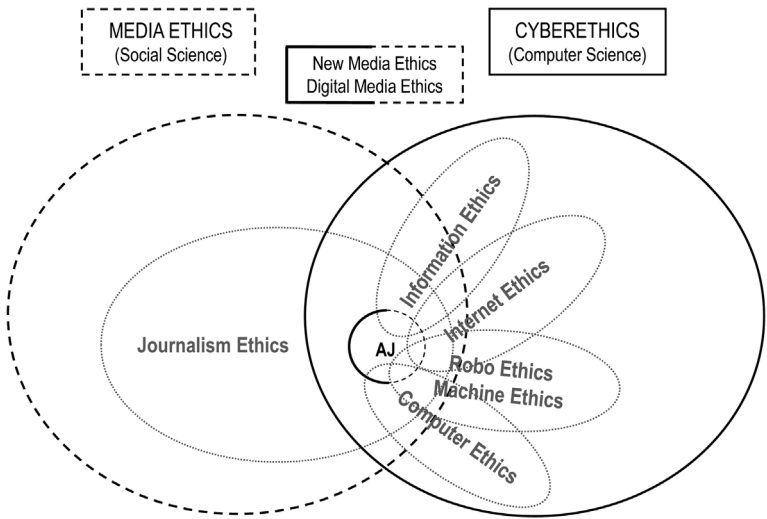


FIGURE 1
Ethical localization of Algorithmic Journalism

(e.g. Groth 1930). Within this field, the social and technological development of journalism shape the discourse which ethical and moral claims are made and what journalism should accomplish for society (Brosda and Schicha 2010).

Throughout this evolution, various values were inscribed into it. At the very beginning, journalism was seen as mere reporting—reproducing facts without editorial interference (Brosda 2010, 259ff.). The major goal was to provide information. Values such as objectivity, neutrality, confirmability, and the role as intermediary were developed. Through social uprisings in the late eighteenth and early nineteenth century, a new interpretation of journalism was established: the expatiating journalism. Journalism should illuminate society, find the truth, criticize the potentate, take a stance, and help to build an opinion. The idea of freedom of opinion and the press were established as well as advocacy for minorities. In the tradition of illumination, investigative journalism was developed, highlighting values of truth and eliminating abuses as well as the protection of sources (Brosda 2010). In the first half of the twentieth century another branch evolved: interpretive journalism (Selgado and Strömbäck 2012)—balancing between reporting and expatiating and trying to compensate differing moral values of objectivity and forming the public opinion.

In the second half of the twentieth century a phase of consolidation of journalism began, i.e. ethically reflected and embedded professional routines evolved on an organizational level. Examples are ethical press codes, e.g. the German “Pressekodex” from 1973, or the International Federation of Journalists’ Declaration of Principles adopted in 1954 (IFJ 2015).

This phase ended through new technological developments by the beginning of the twenty-first century. Kovach and Rosenstiel (2007) state therefore that the traditional “journalism of verification” with its rigorous fact checking is being displaced by deprofessionalized forms of journalism. First, “journalism of assertion” is going back to a mere reporting tradition with little investigation done by the journalist who only enriches existing facts with background information. Parallel, “post-fact” or “after-the-fact checking” journalism evolved (Shirky 2014, 15ff.). Values like scrutiny and correctness therefore suffered. Second, “journalism of affirmation” is primarily providing opinions. Values such as objectivity and advocacy for the weak or the public collide. Third, with “journalism of aggregation,” algorithms are finally introduced. They scan various online resources (news agencies, websites, blogs, PR agencies) for content without fact checking, proofing reliability of sources, or paying for the content. Values such as reliability and correctness are affected, and moral rules such as “You shall not steal” are broken (Debatin 2015).

Although questions of data search and origin (Bradshaw 2014), algorithmic authority (Carlson 2014), algorithmic objectivity (Gillespie 2014), data usage and abuse (Zion and Craig 2014), reasoning (Kraemer, van Overveld, and Peterson 2011; Young and Hermida 2014), transparency (McBride and Rosenstiel 2014a), and accountability (Diakopoulos 2015) were already part of the scientific discourse of algorithmic applications, the ethical challenges of Algorithmic Journalism have not yet been framed and analyzed.

Focusing on the latest research in journalism ethics (McBride and Rosenstiel 2014b; Prinzing et al. 2015), traditional challenges like objective reporting, venality, the separation of fact and fiction, PR or advertising, as well as the correctness of reported facts, respect for privacy and copyright, or the application of appropriate research methods also have to be discussed within the context of an automated content creation and its use and dissemination in professional journalism.

Analytical Framework

This section outlines the analytical framework to uncover the ethical challenges of Algorithmic Journalism on various levels.

As there is a strong tradition of virtue-based ethics at the micro-level in journalism ethics, this line of thinking states that the journalist as an individual has a very strong responsibility for acting morally. He or she should have a diligent, reliable, honest, and trustworthy character and act with discernment. These virtues are often seen as given or to be primed in journalistic education (Hömborg and Klenk 2014).

Common criticism expressed the strong tradition of individual ethics in journalism ethics due to the strong differentiation of the media system and that it is hardly possible to make a single person responsible for a certain outcome. Furthermore, professional journalists may be exposed to external constraints (economic, hierarchical) that

sometimes prevent moral action (see the concept of corporate responsibility in media ethics; Funiok 2011 and 2015; Kraemer, van Overveld, and Peterson 2011). This critique also applies to Algorithmic Journalism as there are multiple actors and different levels of responsibility visible during news production. Therefore, this article identifies ethical challenges for media organizations, professional journalists, service providers, programmers, data providers, and the audience on an organizational, professional, and social sphere.

Even though there have been various attempts at systematizing journalism ethics, e.g. by Pöttker (1999) with his three imperatives of separation, Thomaß's (2003) five ethical principles of journalistic practice, or Haller's (1992) classification along various value structures, the multi-layer system of responsibility introduced by Pürer (1992 referring to Speamann 1977) is on a more abstract level. Besides the individual level, it also takes the media system and the audience into account.

Following Pürer's approach to analyze the ethical challenges of Algorithmic Journalism on different levels, the framework is complemented by the classification of journalism according to Weischenberg, Malik, and Scholl (2006). For an institutional understanding of journalism, it is structured into an organizational, professional, and social sphere. On a social level, journalism fulfills certain tasks by observing parts of society and providing the public with relevant information. This includes fact-based products with a certain reach (Weischenberg, Malik, and Scholl 2006, 346) as well as products of special interest (Rühl 1980, 382). Journalism is traditionally produced on the organizational level within media organizations according to specific rules and routines or by other journalistic actors on a professional level (Hohlfeld 2003, 127; Wolf 2014, 66). It is essential for this theoretical understanding that media organizations are seen as institutions which fulfill specific functions for society (Weischenberg, Malik, and Scholl 2006, 347). The ethical analysis will show that there is an evident shift of functions and challenges within these spheres.

As mentioned above, the levels identified by Pürer (1992) are similar to those of Weischenberg, Malik, and Scholl (2006) and therefore are able to frame the ethical discussion in all its range.

First, the level of individual ethics relying on the morality of the individual journalist is reflected in the professional sphere. It also refers to legal regulations as well as press codes, e.g. the requirement to fulfill the principles of objective, transparent, and accurate reporting as well as appropriate research methods (on the governance of algorithms, see Saurwein, Just, and Latzer 2015). Of course, individual ethics in the professional sphere are shaped and influenced by organizational, societal, and cultural structures, as well as professional and individual principles of journalistic work. In Algorithmic Journalism these principles are embedded within code, with journalists and coders working together to fit the product to individual and organizational ethical standards. We argue that in Algorithmic Journalism (moral) human agency is partly delegated to algorithms. In terms of algorithmic reality construction, Just and Latzer (forthcoming)—referring to Mitcham (2014)—note that “agency is not only imposed, as in traditional mass-media technology, but predominantly delegated, allowing limited algorithmic intentions and autonomy” (forthcoming). In addition, they conclude that this degree of complexity in the cooperation between journalists and algorithmic agents evokes ethical challenges (e.g. low transparency, controllability, accountability) (Just and Latzer, forthcoming). Albeit in alleviated terms, individual ethics are in play as

TABLE 1
Framework for the ethical analysis based on Pürer’s (1992) multi-layer concept of responsibility and spheres of journalism by Weischenberg, Malik, and Scholl (2006)

MORAL THEORIES		
Organizational Sphere	Professional Sphere	Social Sphere
Ethics of the Media System	Individual Ethics	Ethics of the Audience
<ul style="list-style-type: none">- Tensions through Hierarchy- Economic Aims	<ul style="list-style-type: none">- Objectivity- Transparency- Accuracy- Respect for Privacy- Appropriate Methods of Research- Separation of Fact & Fiction	<ul style="list-style-type: none">- Media Education- Control of Demand
FRAMEWORK		

the journalist is in hierarchical control of the final product and is constantly able to adapt the software or change the underlying data-set if ethical issues arise.

Second, the level of the media system extends individual ethics to a specific perspective of tiered responsibilities of legislators, media staff, and media owners. This also reflects the structure of society, politics, economic demands, and hierarchies within media organizations that are mirrored in the organizational sphere.

Third, Pürer ascribes an ethical responsibility to the audience since journalistic production is influenced by public demand. The audience has the power to dismiss specific journalistic products (economic power). This level can be combined with the social sphere. Nevertheless, the different levels of responsibility cannot be clearly separated from each other as the example of press codes show: they reflect standards for working routines of the individual journalist, but they are also incorporated at an organizational level (the corporation or the industry).

For further analysis, this multilevel framework is also embedded in the grand moral theories of deontology, utilitarianism, contractualism, and virtue ethics—as pictured above—as they are helpful in discussing specific ethical problems of Algorithmic Journalism (Table 1).

As journalism has to face ethical challenges arising with new technologies, it is valuable to identify possible areas of conflict from a technological view. The attributes derived from the technical potential of Algorithmic Journalism are therefore linked to the framework of Weischenberg, Malik, and Scholl (2006) and Pürer (1992).

Ethical Challenges Derived from Technology

Algorithmic Journalism is here defined as:

the (semi)-automated process of natural language generation by the selection of electronic data from private or public databases (input), the assignment of relevance of pre-selected or non-selected data characteristics, the processing and structuring of the relevant data sets to a semantic structure (throughput), and the publishing of the final text on an online or offline platform with a certain reach (output). (Dörr 2015, 3)

The starting point for NLG is a database, for example sports, financial, weather, or traffic data. This data can be accessed via cloud or local memory. Then, it is processed according to predefined linguistic and statistical rules to a text in natural language (for an overview, see also Carstensen et al. 2010). The processing is also possible via cloud access or a local copy of the software. Because the result in the form of a text, e.g. the latest match summary of a football game, has to be pre-set in the code, the generation process—the actual “writing”—as well as most of the publication process (output) is almost completely automated without human interference. The final stage of publishing depends on the implementation of the software within content management and online publishing systems of a media organization (on the technical potential, see Dörr 2015).

These changing processes in news production are leading to shifts in journalistic functions and evoke ethical problems which are now addressed (for an overview, see Table 2). Based on the technical premises, we are able to identify upcoming challenges before (input), during (throughput), and after (output) content generation along the spheres of journalism and ethical layers of responsibility.

TABLE 2
Ethical challenges of Algorithmic Journalism

MORAL THEORIES			
	Organizational Sphere	Professional Sphere	Social Sphere
	Ethics of the Media System	Individual Ethics	Ethics of the Audience
Input	DATA <ul style="list-style-type: none">• <i>Data Origin:</i><ul style="list-style-type: none">- Reliability (Accuracy)- Objectivity- Responsibility- Respect for Privacy- Appropriate Methods of Data Collection- Bias• Data Rights / Data Authority• Economic Aims CODE <ul style="list-style-type: none">• <i>Code Structure:</i><ul style="list-style-type: none">- Transparency- Objectivity- Responsibility- Accuracy- Bias- Machine Communication optimization• Code Rights / Code Authority:<ul style="list-style-type: none">- Responsibility• Economic Aims	CODE <ul style="list-style-type: none">• <i>Code Structure:</i><ul style="list-style-type: none">- Transparency- Objectivity- Responsibility- Accuracy- Bias- Machine communication optimization	DATA <ul style="list-style-type: none">• <i>Data Origin:</i><ul style="list-style-type: none">- Data Diversity for Observing Parts of Society
Throughput	• Production authority	• Software / Code Testing & Monitoring	
Output	• Result Authority <ul style="list-style-type: none">• Transparency:<ul style="list-style-type: none">- Disclosure Data Source- Disclosure of Automation- Disclosure of Code• Legal accountability	• Result Monitoring <ul style="list-style-type: none">• Causing Authority	• Transparency & Trust <ul style="list-style-type: none">• Control of Demand• Confirmability of Data and Text Analysis• Observation of Society in only Localized Areas
FRAMEWORK			

First, ethical issues are identified at the input-level concerning data. As structured data is a precondition of NLG and its journalistic use, questions about data origin occur. The data has to be reliable and accurate. This can be evident for e.g. financial reporting, when incorrect share prices are reported based on false data. Software is always able to do erroneous reporting based on poor coding or data sources. As NLG is not yet able to be creative or self-learning (see Reiter 2010), there still is a human element in Algorithmic Journalism in terms of coding:

In a story Aug. 5 about Graham Holdings Co. earnings—generated by Automated Insights using data from Zacks Investment Research—The Associated Press reported incorrectly that Graham Holdings shares had fallen since the start of the year and a year ago. Adjusted for spinoffs, its shares are up about 33 percent this year and 62 percent from a year ago. (Fox News 2015)

Another question is if there are missing data or items in disorder that limit the information value of the data-set. Missing items can lead to biases during content generation as it is only possible to process the information displayed in the database. Does the data meet the prerequisite of objectivity? Is it possible to report the whole story? Where does the data come from? Who collected it? How was it collected?

Traditionally, source protection is of high relevance in journalism. With Algorithmic Journalism, it is questionable whether source protection is possible or even desired as service providers and their journalistic clients should disclose all data sources in terms of data transparency (Ward 2014). Another important question is whether personal and/or privacy rights are infringed when collecting data. Therefore, questions of data origin and quality were once situated at the individual level of responsibility in traditional journalism. Now, this responsibility shifts to the management level of media organization or third parties as they decide which data is used for their product. Issues of data rights and authority also arise. Due to the complexity of programming, NLG is often outsourced to specialized companies like Automated Insights or Retresco. This leads to questions about data processing. Do media organizations have the right to process the data and use the generated texts for commercial use (copyright)? In addition, media organizations in journalism often deal with economic constraints questioning the underlying imperative: collecting data and meeting journalistic values versus processing data for commercialized products?

As Algorithmic Journalism is only possible within programmed semantic structures and pre-set rules, questions about code are also evident at the input level. Does the code reflect journalistic values of media organizations such as transparency, objectivity, responsibility, and accuracy? For example, it took the Associated Press one year to match the software with its own writing style (ASBPE 2015). How difficult is it then to inscribe ethical values for different and more complex products of Algorithmic Journalism? Are media organizations able to understand programming and are they able to disclose the relevance and source assignment to public demand? Furthermore, as the major goal of a coder is to solve a certain problem and to “make it work,” the engineer’s working routines have to be reshaped. Journalistic values thus should be embedded in code (Del Campo et al. 2013). To adapt code for an ethical use in journalism, for example Friedman, Kahn, and Borning (2008) propose a value-sensitive design.

According to the authors (2008, 69), a “value sensitive design is a theoretically grounded approach to the design of technology that accounts for human values in a principled and comprehensive manner throughout the design process.” Here, responsibilities also shift from the individual level to the level of the organization or third parties.

In throughput, media organizations take over production authority. They also have to address questions of authority and transparency like data source disclosure, disclosure of automation and code, as well as legal accountability on the output level (for legal issues of automated content in the United States, see Weeks 2014). So far, the Associated Press tags their texts with a special line that indicates they were generated by Automated Insights using data provided by Zacks (AI 2015). But as Dörr (2015, 14) states, there are other journalistic entities that are not labeling automated content.

Formerly, the option of publishing anonymously was a basic precondition for protecting sources or the journalist from prosecution. Nevertheless, this collides with other values such as transparency and accountability. If the audience should fulfill a moral responsibility (social sphere), it is important whether sources and authors are visible or not. If there is no information on the algorithmic nature of a text and its resources, how should the audience make a decision whether it wants to consume the information and whether it can rely on it? Therefore, information asymmetries arise. Recent empirical findings suggest that the audience cannot fulfill its moral responsibility because it can hardly distinguish unmarked journalistic content created by human journalists from algorithmically created content (e.g. Clerwall 2014; Graefe et al. 2015; Van der Kaa and Krahmer 2014). Furthermore, the audience is not able to decide whether the provided information is correct or has a statistical bias. Since the audience is controlling the demand and shape of Algorithmic Journalism, it decides the topics of observation (see also data diversity for an observation of society as a prerequisite on the input level).

In the professional sphere and due to functional shifts, journalistic actors have to face the same issues as media organizations regarding code structure. In throughput and output, questions of authority arise: Who is monitoring the machine/software/code if errors occur during production? The Associated Press, for example, has stopped monitoring every single generated text for their earnings reports as it is too time consuming (Turi2 2015). This monitoring is also relevant on the output level as results are often published autonomously by the system (ASBPE 2015).

As it is not possible to illustrate the many layers of ethical issues within this article, Table 2 structures the ethical challenges of Algorithmic Journalism and provides a starting point for further research.

Conclusion

With algorithms constructing reality in various life domains and with the gradual delegation of (moral) agency by artificial intelligence (Just and Latzer, forthcoming), the institutionalization of algorithms as content creators evokes various ethical challenges for professional journalism.

First, this article locates these challenges in the overlapping fields of digital media ethics and cyber ethics, with a special focus on journalism ethics (RQ1). Moral theories

of utilitarianism, deontology, contractualism, and virtue ethics are a basis to reflect the impact of Algorithmic Journalism for further research.

In addition, the article provides an ethical framework to identify and discuss these issues. The analysis was framed by Weischenberg, Malik, and Scholl's (2006) understanding of journalism and based on Pürer's (1992) multi-layer system of responsibility that assigns moral responsibility on an organizational, professional/individual, as well as on a social sphere. A technical view of Algorithmic Journalism further helps to identify ethical challenges before, during, and after news production.

As a result of RQ2, a significant shift of responsibility in news production can be observed. With the emergence of Algorithmic Journalism, the human journalist—the individual—is not the major moral agent anymore as other actors, journalistic and non-journalistic, are involved in news production on various levels, e.g. algorithms with delegated agency, media organizations, programmer/service providers of NLG, or data collectors. Thus, the importance of the individual is diminishing, whereas the importance of media organizations and the media system as moral agents is rising.

As non-journalistic actors are able to use NLG and thus have the possibility to reach an audience with a certain scope, a further challenge will be to develop and codify ethical guidelines building an ethical background for non-journalistic actors involved. This also applies to media organizations as they have to develop and adopt ethical codes of conduct for Algorithmic Journalism.

Other challenges caused by hierarchies within the media organization and the editorial department decline through the technical nature and coding of Algorithmic Journalism through external providers. Nevertheless, biases can be an ethical issue while collecting or processing data. Therefore, legal and regulatory questions arise (Darling 2012; Weeks 2014; Saurwein, Just, and Latzer 2015).

The relationship between journalists and coders during news production is an object for further research as well as it is relevant for a better understanding of power structures behind the algorithm.

The role of the audience is also to be examined but it may not be the adequate level for demanding moral responsibility as it can hardly distinguish between human and automated content if it is not labeled.

Finally, it is important to notice that Algorithmic Journalism is also able to overcome the ethical challenges of the conventional media system. It may unburden the human journalist from daily routine work, it may reduce economic pressure and the dictate of quantity. It frees up capacities for in-depth analysis and reporting and thus enables journalists to consider moral demands in journalistic work like checking of multiple sources, reflection, diligence, respecting human dignity, etc.

For future research, it is necessary to extend the discussion of responsibility in journalism under the influence of algorithmic processes and applications. Therefore, journalistic and non-journalistic actors on the organizational, individual/professional, and social/audience level have to develop guidelines along moral and ethical values, e.g. a value-sensitive design for algorithmic applications. This is the only way to meet the same ethical standards for Algorithmic Journalism as for all other professional journalistic products.

ACKNOWLEDGEMENTS

The authors would like to thank Moritz Büchi and Andy Williams as well as the two anonymous reviewers for their helpful and constructive comments.

DISCLOSURE STATEMENT

No potential conflict of interest was reported by the authors.

FUNDING

The research was supported by a research grant of the Swiss National Science Foundation.

REFERENCES

- Al. 2015. "Automated Insights and the Associated Press." <http://automatedinsights.com/ap/>.
- Anderson, Christopher W. 2012. "Towards a Sociology of Computational and Algorithmic Journalism." *New Media and Society* 15 (7): 1005–1021.
- Ananny, Mike. 2015. "Toward an Ethics of Algorithms: Convening, Observation, Probability, and Timeliness." *Science, Technology, & Human Values* 41 (1): 93–117.
- Aristotle. 2003. "The Nicomachean Ethics." In *Virtue Ethics*, edited by Stephen Darwall, 7–50. Malden: Blackwell.
- ASBPE. 2015. "AP's Tom Kent: The Time Has Arrived for Robotics Journalism Ethical Checklist." *American Society of Business Publication Editors*. <http://www.asbpe.org/blog/2015/05/03/aps-tom-kent-the-time-has-arrived-for-robotics-journalism-ethical-checklist/>.
- Bentham, Jeremy. 2015 [1825]. *The Rationale of Reward*. London: Forgotten Books.
- Borenstein, Jason, and Yvette Pearson. 2012. "Robot Caregivers: Ethical Issues across the Human Lifespan." In *Robot Ethics: The Ethical and Social Implications of Robotics*, edited by Patrick Lin, Keith Abney and George A. Bekey, 251–265. Cambridge (MA): MIT Press.
- Bradshaw, Paul. 2014. "Data Journalism." In *Ethics for Digital Journalists: Emerging Best Practices*, edited by Lawrie Zion and David Craig, 202–219. New York: Routledge.
- Brosda, Carsten. 2010. "Journalismus [Journalism]." In *Handbuch Medienethik [Handbook Media Ethics]*, edited by Carsten Brosda and Christian Schicha, 257–277. Wiesbaden: VS Verlag.
- Brosda, Carsten, and Christian Schicha. 2010. "Einleitung [Introduction]." In *Handbuch Medienethik [Handbook Media Ethics]*, edited by Carsten Brosda and Christian Schicha, 9–17. Wiesbaden: VS Verlag.
- Broussard, Meredith. 2014. "Artificial Intelligence for Investigative Reporting. Using an Expert System to Enhance Journalists' Ability to Discover Original Public Affairs Stories." *Digital Journalism* 6 (3): 814–831. doi: 10.1080/21670811.2014.985497.

- Capurro, Rafale. 2007. "Intercultural Information Ethics." In *Localizing the Internet. Ethical Aspects in Intercultural Perspective*, edited by Rafael Capurro, Johannes Frühbauer, and Thomas Hausmanninger, 21–38. Munich: Fink.
- Carlson, Matt. 2014. "The Robotic Reporter. Automated Journalism and the Redefinition of Labor, Compositional Forms, and Journalistic Authority." *Digital Journalism* 3 (3): 416–431. doi: [10.1080/21670811.2014.976412](https://doi.org/10.1080/21670811.2014.976412).
- Carstensen, Kai-Uwe, Christian Ebert, Cornelia Ebert, Susanne Jekat, Ralf Klabunde, and Hagen Langer. 2010. *Computerlinguistik und Sprachtechnologie. Eine Einführung* [Computational Linguistics and Speech Technology. An Introduction]. Heidelberg: Spektrum.
- Clerwall, Christer. 2014. "Enter the Robot Journalist. Users' Perceptions of Automated Content." *Journalism Practice* (8) 5: 519–531.
- Culver, Kathleen Bartzen. 2016. "Disengaged Ethics. Code Development and Journalism's Relationship with 'the Public'." *Journalism Practice*: 1–17. doi: [10.1080/17512786.2015.1121788](https://doi.org/10.1080/17512786.2015.1121788).
- Darling, Kate. 2012. "Extending Legal Rights to Social Robots." *SSRN Online Journal*, 1–18. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2044797.
- Darwall, Stephan. 2003. *Contractarianism/Contractualism*. Malden: Blackwell.
- Davis, Michael, Andrew Kumiega, and Ben Vliet. 2012. "Ethics, Finance, and Automation: A Preliminary Survey of Problems in High Frequency Trading." *Science and Engineering Ethics* 19: 851–874.
- Debatin, Bernhard. 2010. "New Media Ethics." In *Handbuch Medienethik* [Handbook Media Ethics], edited by Carsten Brosda and Christian Schicha, 318–327. Wiesbaden: VS Verlag.
- Debatin, Bernhard. 2015. "Das Ende der journalistischen Ethik? [The End of Ethics in Journalism?]" In *Neuvermessung der Medienethik. Bilanz, Themen und Herausforderungen seit 2000*, edited by Marlis Prinzing, Matthias Rath, Christian Schicha, and Ingrid Stapf, 56–73. Weinheim/Basel: Beltz Juventa.
- Del Campo, Matias, Adam Fure, Wes McGee, Sandra Manninger, and Arthur Flexer. 2013. "Autonomous Tectonics—A Research into Emergent Robotics Construction Methods." In *Rethinking Prototyping: Proceedings of the Design Modelling Symposium Berlin*, edited by Fabian Scheurer, Julien Nembrini, Axel Kilian, and Christoph Gengnagel, 1–13. Berlin.
- Diakopoulos, Nicholas. 2015. "Algorithmic Accountability. Journalistic Investigation of Computational Power Structures." *Digital Journalism* 3 (3): 398–415.
- Dörr, Konstantin. 2015. "Mapping the Field of Algorithmic Journalism." *Digital Journalism*: 1–24. doi:[10.1080/21670811.2015.1096748](https://doi.org/10.1080/21670811.2015.1096748).
- Ess, Charles. 2009. *Digital Media Ethics*. London: Polity Press.
- Fox News. 2015. "Correction: Earns-Graham Holdings Story". <http://www.foxnews.com/us/2015/08/07/correction-earns-graham-holdings-story/>.
- Friedman, Batya, Peter H. Kahn, and Alan Borning. 2008. "Value Sensitive Design and Information Systems." In *The Handbook of Information and Computer Ethics*, edited by Kenneth E. Himma and Herman T. Tavani, 69–101. New Jersey: Wiley.
- Funiok, Rüdiger. 2011. *Medienethik. Verantwortung in der Mediengesellschaft* [Media Ethics. Responsibility in Media Society]. Stuttgart: Kohlhammer.
- Funiok, Rüdiger. 2015. "Hauptthemen und Autoren in der Entwicklung der deutschsprachigen Kommunikations- und Medienethik [Topics and Authors of the Development of

- the German Communications- and Media Ethics]". In *Neuvermessung der Medienethik. Bilanz, Themen und Herausforderungen seit 2000* [Remapping Media Ethics. Results, Topics, and Challenges since 2000], edited by Marlis Prinzing, Matthias Rath, Christian Schicha, and Ingrid Stapf, 20–34. Beltz Juventa: Weinheim/Basel.
- Gillespie, Tarleton. 2014. "The Relevance of Algorithms." In *Media Technologies. Paths Forward in Social Research*, edited by Tarleton Gillespie, Pablo Boczkowski, and Kirsten Foot, 167–194. London: MIT Press.
- Graefe, Andreas, Mario Haim, Bastian Haarmann, and Hans-Bernd Brosius. 2015. Perception of Automated Computer-Generated News: Credibility, Expertise, and Readability. Paper presented at the 11th Dubrovnik Media Days, Dubrovnik.
- Groth, Otto. 1930. *Die Zeitung. Ein System der Zeitungskunde (Journalistik)* [The Newspaper. A System of Journalism]. Mannheim: J. Bensheimer.
- Haller, Michael. 1992. "Die Journalisten und der Ethikbedarf [Journalists and the Need for Ethics]." In *Medien-Ethik [Media Ethics]*, edited by Michael Haller and Helmut Holzhey, 196–211. Opladen: Westdeutscher Verlag.
- Hohlfeld, Ralf. 2003. *Journalismus Und Medienforschung. Theorie, Empirie, Transfer* [Journalism and Media Research. Theory, Empirical Research, Transfer]. Konstanz: UTB.
- Hömborg, Walter and Christian Klenk. 2014. *Die Verantwortung des Journalisten. Individual- und professionsethische Maximen* [The Responsibility of a Journalist. Individual and Professional Ethical Maxims]. Wiesbaden: Springer VS.
- Horner, David S. 2013. *Understanding Media Ethics*. Brighton: SAGE.
- IFJ. 2015. "IFJ Declaration of Principles on the Conduct of Journalists." International Federation of Journalists. <http://www.ifj.org/about-ifj/ifj-code-of-principles/>.
- Johnson, Deborah G. 2009. *Computer Ethics*. London et al.: Pearson.
- Just, Natascha and Michael Latzer. forthcoming. "Governance by Algorithms: Reality Construction by Algorithmic Selection on the Internet". *Media, Culture & Society*: 1–26.
- Kant, Immanuel. 2003 [1797]. "The Metaphysics of Morals." In *Deontology*, edited by Stephen Darwall, 21–27. Malden: Blackwell.
- Kernaghan, Kenneth. 2014. "Digital Dilemmas: Values, Ethics and Information Technology." *Canadian Public Administration* 57 (2): 295–317.
- Kovach, Bill, and Tom Rosenstiel. 2007. *The Elements of Journalism: What Newspeople Should Know and the Public Should Expect*. New York: Three Rivers Press.
- Kraemer, Felicitas, Klees van Overveld, and Martin Peterson. 2011. "Is There an Ethics of Algorithms?" *Ethics and Information Technology* 13 (3): 251–260.
- Langford, Duncan. 2000. *Internet Ethics*. New York: St. Martin's Press.
- Lin, Patrick. 2015. "Why Ethics Matters for Autonomous Cars." In *Autonomes Fahren. Technische, Rechtliche und Gesellschaftliche Aspekte* [Autonomous Driving. Technical, Legal, and societal aspects], edited by Markus Maurer, Christian J. Gerdes, Barbara Lenz, and Hermann Winner, 69–86. Berlin, Heidelberg: Springer.
- Maner, Walter. 1996. "Unique Ethical Problems in Information Technology." *Science and Engineering Ethics* (2) 2: 137–154.
- McBride, Kelly, and Tom Rosenstiel. 2014a. "New Guiding Principles for a New Era of Journalism." In *The New Ethics of Journalism*, edited by Kelly McBride and Tom Rosenstiel, 1–6. Thousand Oaks, CA: CQ Press.
- McBride, Kelly, and Tom Rosenstiel. 2014b. *The New Ethics of Journalism*. Thousand Oaks, CA: CQ Press.

- Meier, Klaus. 2002. "Neue journalistische Formen [New Journalistic Genres]." In *Internet-Journalismus* [Internet Journalism], edited by Klaus Meier, 21–172. Konstanz: UVK.
- Mill, John Stuart. 2006. *Utilitarianism*. Stuttgart: Reclam.
- Mitcham, Carl. 2014. "Agency in Humans and Artifacts: A Contested Discourse." In *The Moral Status of Technical Artefacts*, edited by Peter Kroes and Peter-Paul Verbeek, 11–29. Dordrecht: Springer.
- Napoli, Philip M. 2014. "Automated Media: An Institutional Theory Perspective on Algorithmic Media Production and Consumption." *Communication Theory* (24) 3: 340–360.
- Parry, Roger. 2011. *The Ascent of Media. from Gilgamesh to Google via Gutenberg*. London, Boston: Nicholas Brealey Pub.
- Pavlik, John. 2000. "The Impact of Technology on Journalism." *Journalism Studies* (1) 2: 229–237.
- Pavlik, John. 2013. "Innovation and the Future of Journalism." *Digital Journalism* (1) 2: 181–193.
- Pierce, Margaret A., and John W. Henry. 1996. "Computer Ethics: The Role of Personal, Informal, and Formal Codes." *Journal of Business Ethics* 15 (4): 425–437.
- Pöttker, Horst. 1999. "Berufsethik für Journalisten? Professionelle Trennungsgrundsätze auf dem Prüfstand [Professional Ethics for Journalists? Principles of Separation on the trial]." In *Kommunikations- und Medienethik* [Communications- and Media Ethics]. Interdisziplinäre Perspektiven, edited by Adrian Holderegger, 299–327. Freiburg: Academic Press Fribourg.
- Prinzing, Marlis, Matthias Rath, Christian Schicha, and Ingrid Stapf. 2015. *Neuvermessung der Medienethik. Bilanz, Themen und Herausforderungen seit 2000* [Remapping Media Ethics. Results, Topics, and Challenges since 2000]. Weinheim/Basel: Beltz Juventa.
- Pürer, Heinz. 1992. "Ethik in Journalismus und Massenkommunikation. Versuch einer Theorien-Synopse [Ethics in Journalism and Mass Communication]." *Publizistik* 37: 304–321.
- Reiter, Ehud. 2010. "Natural Language Generation." In *The Handbook of Computational Linguistics and Natural Language Processing*, edited by Alexander Clark, Chris Fox and Shalom Lappin, 574–598. Oxford: Wiley-Blackwell.
- Reiter, Ehud, and Robert Dale. 2000. *Building Natural Language Generation Systems*. Cambridge: Cambridge University Press.
- Ross, William David. 2003 [1930]. "The Right and the Good." In *Deontology*, edited by Stephen Darwall, 55–79. Malden: Blackwell.
- Rühl, Manfred. 1980. *Journalismus und Gesellschaft. Bestandsaufnahme und Theorieentwurf* [Journalism and Society]. Mainz: v. Hase & Koehler.
- Saurwein, Florian, Natascha Just, and Michael Latzer. 2015. "Governance of Algorithms: Options and Limitations." *Info* 17 (6): 35–49.
- Selgado, Susana, and Jesper Strömbäck. 2012. "Interpretive Journalism: A Review of Concepts, Operationalizations and Key Findings." *Journalism* 13 (2): 144–161.
- Shirky, Clay. 2014. "Truth without Scarcity, Ethics without Force." In *The New Ethics of Journalism*, edited by Kelly McBride and Tom Rosenstiel, 9–24. Thousand Oaks: CQ Press.
- Shulman, Carl, Nick Tarleton, and Henrik Jonsson. 2009. "Which Consequentialism? Machine Ethics and Moral Divergence." In *AP-CAP 2009: The Fifth Asia-Pacific Computing and Philosophy Conference*, October 1–2, University of Tokyo, Japan, Proceedings, edited by Carson Reynolds and Alvaro Cassinelli, 23–25. AP-CAP 2009.
- Speamann, Robert. 1977. *Zur Kritik der politischen Utopie* [About the critique of political Utopia]. Stuttgart: Klett.

- Spinello, Richard A. 2011. *Cyberethics. Morality and Law in Cyberspace*. Sudbury: Jones & Bartlett.
- Tavani, Herman H. 2011. *Ethics and Technology. Controversies, Questions, and Strategies for Ethical Computing*, New Jersey: Wiley.
- Thomaß, Barbara. 2003. "Fünf ethische Prinzipien journalistischer Praxis." In *Kommunikations- und Medienethik*, edited by Bernhard Debatin and Rüdiger Funiok, 159–168. Konstanz: UVK.
- Turi2. 2015. "AP Setzt Auf Roboterjournalismus." <http://www.turi2.de/aktuell/ap-setzt-auf-roboterjournalismus/>.
- van Dalen, Arjen. 2012. "The Algorithms behind the Headlines. How Machine-Written News Redefines the Core Skills of Human Journalists." *Journalism Practice* (6) 5-6: 648–658.
- Van der Kaa, Hille, and Emiel Krahmer. 2014. "Journalist versus News Consumer: The Perceived Credibility of Machine Written News." Research Paper Presented at the 2014 Computation + Journalism Symposium, Columbia University, New York, NY. http://compute-cuj.org/cj-2014/cj2014_session4_paper2.pdf.
- Ward, Stephen J. 2006. *The Invention of Journalism Ethics. the Path to Objectivity and beyond*. Montreal: McGill-Queen's University Press.
- Ward, Stephen J. 2014. "The Magical Concept of Transparency." In *Ethics for Digital Journalists: Emerging Best Practices*, edited by Lawrie Zion and David Craig, 45–58. New York: Routledge.
- Weeks, Lynn. 2014. "Media Law and Copyright Implications of Automated Journalism." In *Journal of Intellectual Property and Entertainment Law* 4 (1): 67–94.
- Weischenberg, Siegfried, Maja Malik, and Armin Scholl. 2006. "Journalismus in Deutschland 2005. Zentrale Befunde der aktuellen Repräsentativbefragung deutscher Journalisten [Journalism in Germany 2005. Central Findings of a Representative Survey of German Journalists]." *Media Perspektiven* 7: 346–361.
- Wolf, Cornelia. 2014. *Mobiler Journalismus. Angebote, Produktionsroutinen und redaktionelle Strategien Deutscher Print- und Rundfunkredaktionen* [Mobile Journalism. Services, Production Routines and Strategies of German Editorial Offices in Print and Audio]. Baden-Baden: Nomos.
- Young, Mary L., and Alfred Hermida. 2014. "From Mr. and Mrs. Outlier to Central Tendencies." *Digital Journalism* (online first), doi: [10.1080/21670811.2014.976409](https://doi.org/10.1080/21670811.2014.976409).
- Zion, Lawrie, and David Craig. 2014. *Ethics for Digital Journalists. Emerging Best Practices*. New York: Routledge.

Konstantin Nicholas Dörr (author to whom correspondence should be addressed),
Institute of Mass Communication and Media Research, Media Change & Innovation Division,
University of Zurich, Switzerland; E-mail: k.doerr@ipmz.uzh.ch
Katharina Hollnbuchner, Institute of Mass Communication and Media Research,
Media Change & Innovation Division, University of Zurich, Switzerland; E-mail:
k.hollnbuchner@ipmz.uzh.ch

VII. Abbildungs- und Tabellenverzeichnis

Abbildungen

Abbildung 1: Verortung und Einfluss der automatisierten Textproduktion im Journalismus in Anlehnung an Weischenberg (1992; et al. 2006), Esser (1998: 27) und Latzer (2013)	28
Abbildung 2: Einordnung von AJ in Marktphasen in Anlehnung an Heuss (1965) und Latzer et al. (2014)	47

Tabellen

Tabelle 1: Übersicht über Forschungsfragen, den wissenschaftlichen Beitrag und Kontext-Ebenen als Bezugspunkte der Fragestellungen	31
Tabelle 2: Überblick Autorenleistung, Artikel III	211
Tabelle 3: Überblick Autorenleistung, Artikel IV	211
Tabelle 4: Überblick Autorenleistung, Artikel V	212

VIII. Literaturverzeichnis

- Acemoglu, D., & Restrepo, P. 2016. The Race between Man and Machine: Implications of Technology for Growth, Factor Shares and Employment. *National Bureau of Economic Research*, Working Paper, 22252. <http://www.nber.org/papers/w22252>.
- AI. 2015a. The Future of Finance is automated. <http://automatedinsights.com/ap/>.
- AI. 2015b. Robot Reporters: How the AP embraced Data Automation | Strata + Hadoop World. <https://www.youtube.com/watch?v=IBzXiMiwQQs>.
- AI. 2015c. When Robots write the News, what will Humans do? | SXSW Interactive. <https://www.youtube.com/watch?v=0LqLHAAtLE-c>.
- Aitamurto, T., & Lewis, S.C. 2013. Open Innovation in Digital Journalism: Examining the Impact of open APIs at four News Organizations. *New Media & Society*, 15(2), 314–331.
- Altmeppen, K.-D. 2000. Entscheidungen und Koordinationen. Dimensionen journalistischen Handelns. In *Theorien des Journalismus. Ein diskursives Handbuch*, herausgegeben von Löffelholz, M., 293–310. Opladen: Westdeutscher Verlag.
- Altmeppen, K.-D. 2003a. Medienökonomie im Internet-Zeitalter. Problemorientierung und Entwicklungspfade. In *Die neue Kommunikationswissenschaft. Theorien, Themen und Berufsfelder im Internet-Zeitalter. Eine Einführung*, herausgegeben von Löffelholz, M., & Quandt, T., 215–234. Wiesbaden: Westdeutscher Verlag.
- Altmeppen, K.-D. 2003b. Ist der Journalismus strukturell qualitätsfähig? Der Stellenwert journalistischer Organisationen, journalistischer Produkte und journalistischer Medien für die Qualität. In *Qualität im Journalismus*, herausgegeben von Bucher, H.J., & Altmeppen, K.-D., 113–128. Wiesbaden: Westdeutscher Verlag.
- Ananny, M. 2016. Toward an Ethics of Algorithms: Convening, Observation, Probability, and Timeliness. *Science, Technology, & Human Values*, 41(1), 93–117.
- Anderson, C.W. 2011a. Deliberative, agonistic, and algorithmic Audiences: Journalism's Vision of its Public in an Age of Audience Transparency. *International Journal of Communication*, 5, 529–547.
- Anderson, C.W. 2011b. Understanding the Role played by Algorithms and Computational Practices in the Collection, Evaluation, Presentation, and Dissemination of journalistic Evidence. *Whitepaper* für das «Berlin Symposium on the Internet and Society», Berlin, Deutschland, 25. bis 28. Oktober. <http://www.hiig.de/wp-content/uploads/2012/04/Intermediaries-in-Public-Communication-Algorithm-and-Journalism-Paper.pdf>.
- Anderson, C.W. 2012. Towards a Sociology of Computational and Algorithmic Journalism. *New Media & Society*, 15(7), 1005–1021.
- Anderson, C.W. 2013. What Aggregators do: Towards a networked Concept of journalistic Expertise in the Digital Age. *Journalism*, 14(8), 1008–1023.
- AP. 2015. Private Equity Firm Vista buying Automated Insights. <http://www.ap.org/Content/AP-In-The-News/2015/Private-equity-firm-Vista-buying-Automated-Insights>.
- Aristotle. 2003. From The Nicomachean Ethics. In *Virtue Ethics*, herausgegeben von Darwall, S., 7–50. Malden: Wiley-Blackwell.
- Arnold, K. 2008. Qualität im Journalismus. Ein integratives Konzept. *Publizistik*, 53(4), 488–508.

- Arntz, M., Gregory, T., & Zierahn, U. 2016. The Risk of Automation for Jobs in OECD Countries: A comparative Analysis. *OECD Social, Employment and Migration Working Papers*, 189, Paris: OECD Publishing.
- Arria, N.L.G. 2015. Termination of Contract and Trading Update. *Email from Stuart Rogers, CEO Arria NLG*, 30(4).
- ASBPE (American Society of Business Publication Editors). 2015. AP's Tom Kent: The Time has arrived for Robotics Journalism ethical Checklist. <http://www.asbpe.org/blog/2015/05/03/aps-tom-kent-the-time-has-arrived-for-robotics-journalism-ethical-checklist/>.
- Atteslander, P. 2006. *Methoden der empirischen Sozialforschung*. Berlin: ES Verlag.
- Autor, D.H. 2015. Why are there still so many Jobs? The History and Future of Workplace Automation. *Journal of Economic Perspectives*, 29(3), 3–30.
- AX Semantics. 2017. Feature Preview: Using NLP for NLG. <https://www.youtube.com/watch?v=aVXRhpy3Ook>.
- Bain, J.S. 1956. *Barriers to new Competition. Their Character and Consequences in Manufacturing Industries*. Cambridge: Harvard University Press.
- Bateman, J. 2010. Angewandte natürlichsprachliche Generierungs- und Auskunftssysteme. In *Computerlinguistik und Sprachtechnologie. Eine Einführung*, herausgegeben von Carstensen, K-U., Ebert, C., Ebert C., Jekat, S., Klabunde, R., & Langer, H., 633–641. Heidelberg: Spektrum.
- Baum, A. 1994. *Journalistisches Handeln. Eine kommunikationstheoretisch begründete Kritik der Journalismusforschung*. Opladen: Westdeutscher Verlag.
- Beck, K. 2013. *Kommunikationswissenschaft*. Konstanz: UVK.
- Beer, D. 2009. Power through the Algorithm? Participatory Web Cultures and the technological Unconscious. *New Media & Society*, 11(6), 985–1002.
- Bell, E. 2012. Journalism by Numbers. http://www.cjr.org/cover_story/journalism_by_numbers.php.
- Bentham, J. 2015. *The Rationale of Reward*. London: Forgotten Books.
- Berger, M., Haim, M., Graefe, A., Brosius, H.B., & Hess, T. 2015. Aktuelles Stichwort: Computational Journalism. *Medienwirtschaft*, 12(1), 20–23.
- Berners-Lee, T., Hendler, J., & Lassila, O. 2001. The Semantic Web. *Scientific American*, 284(5), 24–30.
- Bieger, T., & Reinhold, S. 2011. Das wertbasierte Geschäftsmodell. Ein aktualisierter Strukturierungsansatz. In *Innovative Geschäftsmodelle*, herausgegeben von Bieger, T., zu Knyphausen-Aufseß, D., & Krys, C., 13–70. Berlin: Springer.
- Binsted, K., & Ritchie, G. 1997. Computational Rules for generating Punning Riddles. *Humor. International Journal of Humor Research*, 10(1), 25–76.
- Blöbaum, B. 2000. Strukturwandel des Journalismus. Strukturwandel von Öffentlichkeit. In *Zerfall der Öffentlichkeit?*, herausgegeben von Jarren, O., Imhof, K., & Blum, R., 135–147. Wiesbaden: Westdeutscher Verlag.
- Blöbaum, B. 2005. Wandel und Journalismus. Vorschlag für einen analytischen Rahmen. In *Journalismus und Wandel. Analysedimensionen, Konzepte, Fallstudien*, herausgegeben

- von Behmer, M., Blöbaum, B., Scholl, A., & Stöber, R., 41–60. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Blöbaum, B., Nölleke, D., & Scheu, A.M. 2016. Das Experteninterview in der Kommunikationswissenschaft. In *Handbuch nicht standardisierte Methoden in der Kommunikationswissenschaft*, herausgegeben von Auerbeck-Lietz, S., & Meyen, M., 175–190. Wiesbaden: Springer.
- Bogner, A., Littig, B., & Menz, W. 2014. *Interviews mit Experten. Eine praxisorientierte Einführung*. Wiesbaden: Springer VS.
- Borchardt, A., & Göthlich, S.E. 2009. Erkenntnisgewinnung durch Fallstudien. In *Methodik der empirischen Forschung*, herausgegeben von Albers, S., Klapper, D., Konradt, U., Walter, A., & Wolf, J., 33–48. Wiesbaden: Gabler.
- Borenstein, J., & Pearson, Y. 2012. Robot Caregivers: Ethical Issues across the Human Lifespan. In *Robot Ethics: The ethical and social Implications of Robotics*, herausgegeben von Lin, P., Abney, K., & Bekey, G.A., 251–265. Cambridge (MA): MIT Press.
- Boyd, D., Crawford, K. 2013. Big Data als kulturelles, technologisches und wissenschaftliches Phänomen. Sechs Provokationen. In *Big Data. Das neue Versprechen der Allwissenheit*, herausgegeben von Geiselberger, H., & Moorstedt T., 187–218. Berlin: Suhrkamp.
- Bradshaw, P. 2014. Data Journalism. In *Ethics for Digital Journalists: Emerging Best Practices*, herausgegeben von Zion, L., & Craig, D., 202–219. New York: Routledge.
- Brosda, C. 2010. Journalismus. In *Handbuch Medienethik*, herausgegeben von Brosda, C., & Schicha, C., 257–277. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Brosda, C., & Schicha, C. 2010. Einleitung. In *Handbuch Medienethik*, herausgegeben von Brosda, C. & Schicha, C., 9–17. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Broussard, M. 2015. Artificial Intelligence for Investigative Reporting. Using an Expert System to enhance Journalists' Ability to discover original Public Affairs Stories. *Digital Journalism*, 3(6), 814–831.
- Brüsemeister, T. 2008. *Bildungssoziologie: Einführung in Perspektiven und Probleme*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Bucher, T. 2012. Want to be on the Top? Algorithmic Power and the Threat of Invisibility on Facebook. *New Media & Society*, 14(7), 1164–1180.
- Callaway, C.B., & Lester, J.C. 2002. Narrative Prose Generation. *Artificial Intelligence*, 139(2), 213–252.
- Capurro, R. 2007. Intercultural Information Ethics. In *Localizing the Internet. Ethical Aspects in Intercultural Perspective*, herausgegeben von Capurro, R., Frühbauer, J., & Hausmanning, T., 21–38. Munich: Fink.
- Carlson, M. 2014. The Robotic Reporter. Automated Journalism and the Redefinition of Labor, Compositional Forms, and Journalistic Authority. *Digital Journalism*, 3(3), 416–431.
- Carstensen, K.-U. 2012. *Sprachtechnologie. Ein Überblick*. <http://www.kai-uwe-carstensen.de/Publikationen/Sprachtechnologie.pdf>.
- Carstensen, K.-U., Ebert, C., Ebert, C., Jekat, S., Klabunde, R., & Langer, H. 2010. *Computerlinguistik und Sprachtechnologie. Eine Einführung*. Heidelberg: Spektrum.

- Cleary, J., Cochie, M. 2011. Core Skill Set remains same in Newspaper Job Ads. *Newspaper Research Journal*, 32(4), 68–82.
- Clerwall, C. 2014. Enter the Robot Journalist. Users' Perceptions of automated Content. *Journalism Practice*, (8)5, 519–531.
- Coddington, M. 2015. Clarifying Journalism's quantitative Turn: A Typology for Evaluating Data Journalism, Computational Journalism, and Computer-Assisted Reporting. *Digital Journalism*, 3(3), 331–348.
- Collins, H.M., & Kusch, M. 1998. *The Shape of Actions. What Humans and Machines can do*. Cambridge (MA): MIT Press.
- Cormen, T.H., Leiserson, C.E., Rivest, R.L., & Stein, C. 2009. *Introduction to Algorithms*. Cambridge (MA): MIT Press.
- Cox, M. 2000. The Development of Computer-Assisted Reporting. Vortrag vor der *Association for Education in Journalism and Mass Communication*, 17. bis 18. März, University of North Carolina, USA. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.631m.6220&rep=rep1&type=pdf>.
- Creech, B., & Mendelson, A.L. 2015. Imagining the Journalist of the Future: Technological Visions of Journalism Education and Newswork. *The Communication Review*, 18(2), 142–165.
- Creswell, J.W. 1994. *Research Design: Qualitative and quantitative Approaches*. Thousand Oaks (CA): Sage.
- Culver, K.B. 2016. Disengaged Ethics. Code Development and Journalism's Relationship with 'The Public'. *Journalism Practice* [online first]. doi: [10.1080/17512786.2015.1121788](https://doi.org/10.1080/17512786.2015.1121788)
- Czygan, M., & Kallfaß, H.H. 2003. Medien und Wettbewerbstheorie. In *Medien und Ökonomie. Grundlagen der Medienökonomie: Kommunikations- und Medienwissenschaft, Wirtschaftswissenschaft*, herausgegeben von Altmeppen, K.-D., & Karmasin, M., 283–304. Wiesbaden: Westdeutscher Verlag.
- Darling, K. 2012. Extending Legal Protection to Social Robots: The Effects of Anthropomorphism, Empathy, and Violent Behavior towards Robotic Objects. *We Robot Conference 2012*, University of Miami. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2044797.
- Darwall, S. 2003. *Contractarianism/Contractualism*. Malden: Blackwell.
- Davis, M., Kumiega, A., & Van Vliet, B. 2012. Ethics, Finance, and Automation: A Preliminary Survey of Problems in High Frequency Trading. *Science and Engineering Ethics*, 19(3), 851–874.
- Debatin, B. 1998. Verantwortung im Medienhandeln. Medienethische und handlungstheoretische Überlegungen zum Verhältnis von Freiheit und Verantwortung in der Massenkommunikation. In *Freiheit und Medien. Beiträge zur Medienethik, Band 4*, herausgegeben von Wunden, W., 113–130. Frankfurt a.M.: Abt. Verlag.
- Debatin, B. 2010. New Media Ethics. In *Handbuch Medienethik*, herausgegeben von Brosda, C., & Schicha, C., 318–327. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Debatin, B. 2015. Das Ende der journalistischen Ethik? In *Neuvermessung der Medienethik. Bilanz, Themen und Herausforderungen seit 2000*, herausgegeben von Prinzing, M., Rath, M., Schicha, C., & Stapf, I., 56–73. Weinheim/Basel: Beltz Juventa.

- Debatin, B., & Funiok, R. 2003. *Kommunikations- und Medienethik*. Konstanz: UVK.
- Decker, M., Fleischer, T., Schippl, J., & Weinberger, N. 2012. Zukünftige Themen der Innovations- und Technikanalyse. *KIT Scientific Reports*, 7605.
- Del Campo, M., Fure, A., McGee, W., Manninger, S., & Flexer, A. 2013. Autonomous Tectonics. A Research into emergent Robotics Construction Methods. In *Rethinking Prototyping: Proceedings of the Design Modelling Symposium Berlin*, herausgegeben von Gengnagel, C., Kilian, A., Nembrini, J., & Scheurer, F., 1–13. Berlin.
- Denzin, N.K. 1970. *The Research Act*. Chicago (IL): Aldine.
- Denzin, N.K. 1989. *The Research Act*. Englewood Cliffs (NJ): Prentice Hall.
- Dernbach, B. 2016. Ausbildung für Journalismus. In *Handbuch Journalismustheorien*, herausgegeben von Löffelholz, M., & Rothenberger, L., 475–487. Wiesbaden: Springer VS.
- Deutscher Fachjournalisten-Verband (DFJV). 2016. *Journalistische Genres*. Konstanz: UVK.
- Deuze, M. 2007. *Media Work*. Cambridge (UK): Polity Press.
- Deuze, M., Neuberger, C., & Paulussen, S. 2004. Journalism Education and Online Journalists in Belgium, Germany, and the Netherlands. *Journalism Studies*, 5(1), 19–29.
- Diakopoulos, N. 2011. A Functional Roadmap for Computational Journalism. http://www.nickdiakopoulos.com/wp-content/uploads/2007/05/CJ_Whitepaper_Diakopoulos.pdf.
- Diakopoulos, N. 2015. Algorithmic Accountability. Journalistic Investigation of Computational Power Structures. *Digital Journalism*, 3(3), 398–415.
- Diakopoulos, N., & Koliska, M. 2016. Algorithmic Transparency in the News Media. *Digital Journalism* [online first]. doi: [10.1080/21670811.2016.1208053](https://doi.org/10.1080/21670811.2016.1208053)
- Djerf-Pierre, M., Ghersetti, M., & Hedman, U. 2016. Appropriating Social Media: The changing Uses of Social Media among Journalists across Time. *Digital Journalism*, 4(6), 849–860.
- Dogruel, L. 2013. *Eine kommunikationswissenschaftliche Konzeption von Medieninnovationen*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Dolata, U. 2011. *Wandel durch Technik: Eine Theorie soziotechnischer Transformation*. Frankfurt a.M.: Campus.
- Dolata, U., & Werle, R. 2007. Bringing Technology back in. In *Gesellschaft und die Macht der Technik. Sozioökonomischer und institutioneller Wandel durch Technisierung*, herausgegeben von Dolata, U., & Werle, R., 15–43. Frankfurt a.M.: Campus.
- Domingo, D., & Paterson, C. 2011. *Making Online News: Newsroom Ethnographies in the Second Decade of Internet Journalism*. New York: Peter Lang.
- Donsbach, W. 1987. Journalismusforschung in der Bundesrepublik: Offene Fragen trotz Forschungsboom. In *Zwischenbilanz der Journalistenausbildung*, herausgegeben von Wilke, J., 105–144. München: Ölschlager.
- Dörr, K. 2016. Algorithmen, Big Data und ihre Rolle im Computational Journalism. In *Journalismusforschung. Stand und Perspektiven, 2., aktualisierte und erweiterte Auflage.*, herausgegeben von Meier, K., & Neuberger, C., 245–264. Baden-Baden: Nomos.
- Dörr, K. 2016. Mapping the Field of Algorithmic Journalism. *Digital Journalism*, 4(6), 700–722.

- Dörr, K., & Hollnbuchner, K. 2016. Ethical Challenges of Algorithmic Journalism. *Digital Journalism* [online first]. doi: [10.1080/21670811.2016.1167612](https://doi.org/10.1080/21670811.2016.1167612)
- Dörr, K., Köberer, N., & Haim, M. 2017. Normative Qualitätsansprüche an algorithmischen Journalismus. In *Gesellschaft ohne Diskurs? Digitaler Wandel und Journalismus aus medienethischer Perspektive. Kommunikations- und Medienethik, 5*, herausgegeben von Stapf, I., Prinzing, M., & Filipović, A., 121–131. Baden-Baden: Nomos.
- Dulinski, U. 2003. *Sensationsjournalismus in Deutschland*. Konstanz: UVK.
- Eilam, E. 2005. *Reversing: Secrets of Reverse Engineering*. Indianapolis (IN): Wiley.
- Esposito, E. 1993. Der Computer als Medium und Maschine. *Zeitschrift für Soziologie*, 22(5), 338–354.
- Ess, C. 2009. *Digital Media Ethics*. London: Polity Press.
- Esser, F. 1998. *Die Kräfte hinter den Schlagzeilen. Englischer und deutscher Journalismus im Vergleich*. Freiburg i.B.: Alber.
- Ettema, J.E., Whitney, D.C., & Wackman, D.B. 1987. Professional Mass Communicators. In *Handbook of Communication Science*, herausgegeben von Berger, C.R., & Chaffee, S.H., 747–780. Beverley Hills (CA): Sage.
- Fabris, H.H. 2000. Vielfältige Qualität. Theoretische Ansätze und Perspektiven der Diskussion um Qualität im Journalismus. In *Theorien des Journalismus*, herausgegeben von Löffelholz, M., 363–374. Wiesbaden: Springer.
- Fabris, H.H., & Renger, R. 2003. Vom Ethik- zum Qualitätsdiskurs. In *Qualität im Journalismus*, herausgegeben von Bucher, H.-J., & Altmeppen, K.-D., 79–91. Wiesbaden: Westdeutscher Verlag.
- Flew, T., Spurgeon, C., Daniel, A., & Swift, A. 2012. The Promise of Computational Journalism. *Journalism Practice*, 6(2), 157–171.
- Flick, U. 2008. *Triangulation. Eine Einführung*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Fög (Forschungsinstitut Öffentlichkeit und Gesellschaft - Universität Zürich). 2015. *Jahrbuch Qualität der Medien 2015 Schweiz*. Basel: Schwabe.
- Fox News. 2015. Correction: Earnings-Graham Holdings Story. <http://www.foxnews.com/us/2015/08/07/correction-earnings-graham-holdings-story/>.
- Frey, B.C., & Osborne, M.A. 2013. The Future of Employment: How susceptible are Jobs to Computerisation? *Working Paper*, Universität Oxford (UK). http://www.oxfordmartin.ox.ac.uk/downloads/academic/The_Future_of_Employment.pdf.
- Friedman, B., Kahn, P.H., & Borning, A. 2008. Value Sensitive Design and Information Systems. In *The Handbook of Information and Computer Ethics*, herausgegeben von Himma, K.E., & Tavani, H.T., 69–101. Hoboken (NJ): Wiley & Sons.
- Fuchs, W., Klima, R., Lautmann, R., & Wienold, H. 1978. *Lexikon der Soziologie*. Reinbek.
- Funiok, R. 2007. *Medienethik. Verantwortung in der Mediengesellschaft*. Stuttgart: Kohlhammer.
- Funiok, R. 2011. *Medienethik. Verantwortung in der Mediengesellschaft. 2., aktualisierte und überarbeitete Auflage*. Stuttgart: Kohlhammer.
- Funiok, R. 2015. Hauptthemen und Autoren in der Entwicklung der deutschsprachigen Kommunikations- und Medienethik. In *Neuvermessung der Medienethik. Bilanz, Themen*

- und Herausforderungen seit 2000, herausgegeben von Prinzing, M., Rath, M., Schicha, C., & Stapf, I., 20–34. Weinheim/Basel: Beltz Juventa.
- Gehrke, G., & Hohlfeld, R. 1995. *Wege zur Theorie des Rundfunkwandels: Fernsehorganisationen zwischen publizistischen Zielvorstellungen und systematischem Eigensinn*. Opladen: Westdeutscher Verlag.
- Gillespie, T. 2014. The Relevance of Algorithms. In *Media Technologies: Essays on Communication, Materiality, and Society*, herausgegeben von Gillespie, T., Boczkowski, P., & Foot, K., 167–194. London: MIT Press.
- Godulla, A. 2017. *Öffentliche Kommunikation im digitalen Zeitalter: Grundlagen und Perspektiven einer integrativen Modellbildung*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Goldberg, E., Driedger, N., & Kittredge, R. 1994. Using Natural-Language Processing to produce Weather Forecasts. *IEEE Expert*, 9(2), 45–53.
- Graefe, A. 2016. Guide to Automated Journalism. <https://towcenter.gitbooks.io/guide-to-automated-journalism/content/>.
- Graefe, A., Haim, M., Haarmann, B., & Brosius, H.-B. 2015. Perception of automated computer-generated News: Credibility, Expertise, and Readability. *Vortrag auf den Dubrovnik Media Days*, 30. bis 31. Oktober, Dubrovnik, Kroatien.
- Graefe, A., Haim, M., Haarmann, B., & Brosius, H.-B. 2016. Readers' Perception of Computer-Generated News: Credibility, Expertise, and Readability. *Journalism* [online first]. doi: [10.1177/1464884916641269](https://doi.org/10.1177/1464884916641269)
- Gray, J., Bounegru, L., & Chambers, L. 2012. *The Data Journalism Handbook. How Journalists Can Use Data to Improve the News*. Sebastopol: O'Reilly.
- Groth, O. 1930. *Die Zeitung. Ein System der Zeitungskunde (Journalistik)*. Mannheim: J. Bensheimer.
- Groth, O. 1960. *Die unerkannte Kulturmacht. Grundlagen der Zeitungswissenschaft (Periodik)*. Berlin: Walter de Gruyter.
- Grunwald, A. 2010. *Technikfolgenabschätzung: Eine Einführung*. Berlin: Edition Sigma.
- Gynnild, A. 2014. Journalism Innovation leads to Innovation Journalism: The Impact of Computational Exploration on changing Mindsets. *Journalism*, 15(6), 713–730.
- Haas, H., & Wallner, C. 2008. Transnational vergleichende Mediensystemforschung: Das erweiterte SCP-Modell und seine Anwendung. In *Medien & Kommunikationsforschung im Vergleich. Grundlagen, Gegenstandsbereiche, Verfahrensweisen*, herausgegeben von Melischek, G., Seethaler, J., & Wilke, J., 83–98. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Häder, M. 2010. *Empirische Sozialforschung. Eine Einführung*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Haim, M., & Dörr, K. 2016. Schema F. Zur Institutionalisierung und normativen Qualität von algorithmischem Journalismus. Vortrag auf der Jahrestagung der DGPK, Leipzig.
- Haim, M., & Graefe, A. 2016. Automated News: Better than expected? Vortrag auf der Jahrestagung der ICA, Fukuoka/Japan.
- Haller, M. 1992. Die Journalisten und der Ethikbedarf. In *Medien-Ethik*, herausgegeben von Haller, M., & Holzhey, H., 196–211. Opladen: Westdeutscher Verlag.

- Haller, M. 2010. Ethik und Qualität. In *Handbuch Medienethik*, herausgegeben von Brosda, C., & Schicha, C., 348–361. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Hallin, D. 2000. Commercialism and the Professionalism in the American News Media. In *Mass Media and Society*, herausgegeben von Curran, J., & Gurevitch, M., 218–237. London: Arnold.
- Hamilton, J.T., & Turner, F. 2009. Accountability through Algorithm. Developing the Field of Computational Journalism. A Report from developing the Field of Computational Journalism, Center for Advanced Study in the Behavioral Sciences Summer Workshop, 27. bis 31. Juli. <http://web.stanford.edu/~fturner/Hamilton%20Turner%20Acc%20by%20Alg%20Final.pdf>.
- Hanitzsch, T. 2009. Zur Wahrnehmung von Einflüssen im Journalismus. Komparative Befunde aus 17 Ländern. *Medien & Kommunikationswissenschaft*, 57(2), 153–173.
- Hanitzsch, T., & Engesser, S. 2014. Journalismusforschung als Integrationsdisziplin. In *Kommunikationswissenschaft als Integrationsdisziplin*, herausgegeben von Karmasin, M., Rath, M., & Thomaß, B., 137–157. Wiesbaden: Springer Fachmedien.
- Harrison, C., & Caglayan, A. 1997. *Agent Sourcebook: A complete Guide to Desktop, Internet, and Intranet Agents*. New York: John Wiley & Sons.
- Häußling, R. 2014. *Techniksoziologie*. Baden-Baden: Nomos.
- Henry, N. 2007. *American Carnival: Journalism under Siege in an Age of New Media*. Berkeley (CA): University of California Press.
- Heuss, E. 1965. *Allgemeine Markttheorie*. Tübingen: Mohr.
- Hoffmann, D. 2005. Experteninterview. In *Qualitative Medienforschung. Eine Einführung*, herausgegeben von Mikos, L., & Wegener, C., 268–278. Konstanz: UVK.
- Hohlfeld, R. 2003. *Journalismus und Medienforschung. Theorie, Empirie, Transfer*. Konstanz: UVK.
- Hohlfeld, R. 2005. Der missachtete Leser revisited. Zum Wandel von Publikumsbild und Publikumsorientierung im Journalismus. In *Journalismus und Wandel. Analysedimensionen, Konzepte, Fallstudien*, herausgegeben von Blöbaum, B., & Scholl, A., 195–224. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Hohlfeld, R. 2010. Publizistische Qualität in neuen Öffentlichkeiten. Crossmedia als Herausforderung für die Verbindung einer Theorie publizistischer Qualität mit einer Theorie publizistischer Innovation. In *Crossmedia – Wer bleibt auf der Strecke?*, herausgegeben von Hohlfeld, R., Müller, P., Richter, A., & Zacher, F., 20–36. Berlin: LIT.
- Hohlfeld, R. 2016. Journalistische Beobachtungen des Publikums. In *Journalismusforschung. Stand und Perspektiven*, herausgegeben von Meier, K., & Neuberger, C., 263–286. Baden-Baden: Nomos.
- Hohlfeld, R., Müller, P., Richter, A., & Zacher, F. 2010. *Crossmedia – Wer bleibt auf der Strecke?* Berlin: LIT.
- Hömberg, W., & Klenk, C. 2014. *Die Verantwortung des Journalisten. Individual- und professionsethische Maximen*. Wiesbaden: Springer VS.
- Horner, D.S. 2013. *Understanding Media Ethics*. Brighton: SAGE.
- IFJ (International Federation of Journalists). 2015. IFJ Declaration of Principles on the Conduct of Journalists. <http://www.ifj.org/about-ifj/ifj-code-of-principles/>.

- Jarren, O. 2008. Massenmedien als Intermediäre. Zur anhaltenden Relevanz der Massenmedien für die öffentliche Kommunikation. *Medien & Kommunikationswissenschaft*, 56(3-4), 329–346.
- Johnson, D.G. 2009. *Computer Ethics*. London: Pearson.
- Jones, K.S. 2001. Natural Language Processing. A historical Review. *Working Paper*, Universität Cambridge (UK). <http://www.cl.cam.ac.uk/archive/ksj21/histdw4.pdf>.
- Jungnickel, K. 2011. Nachrichtenqualität aus Nutzersicht. Ein Vergleich zwischen Leserurteilen und wissenschaftlich-normativen Qualitätsansprüchen. *Medien & Kommunikationswissenschaft*, 59(3), 360–378.
- Just, N., & Latzer, M. 2016. Governance by Algorithms: Reality Construction by Algorithmic Selection on the Internet. *Media, Culture & Society* [online first]. doi: [10.1177/0163443716643157](https://doi.org/10.1177/0163443716643157)
- Kant, I. 1788. *Kritik der praktischen Vernunft*. Riga: Hartknoch.
- Kant, I. 2003. The Metaphysics of Morals. In *Deontology*, herausgegeben von Darwall, S., 21–27. Malden: Blackwell.
- Karmasin, M. 2003. Was ist neu an der neuen Kommunikationswissenschaft? In *Die neue Kommunikationswissenschaft. Theorien, Themen und Berufsfelder im Internet-Zeitalter. Eine Einführung*, herausgegeben von Löffelholz, M., & Quandt, T., 49–57. Wiesbaden: Westdeutscher Verlag.
- Karmasin, M. 2005. *Journalismus: Beruf ohne Moral? Von der Berufung zur Profession*. Wien: Facultas.
- Katzenbach, C. 2012. Technologies as Institutions. In *Trends in Communication Policy Research*, herausgegeben von Just, N., & Puppis, M., 117–137. Bristol: Intellect.
- Kernaghan, K. 2014. Digital Dilemmas: Values, Ethics and Information Technology. *Canadian Public Administration*, 57(2), 295–317.
- Kiefer, M.L. 2001. *Medienökonomik*. München: Oldenbourg.
- Kiefer, M.L. 2010. *Journalismus und Medien als Institutionen*. Konstanz: UVK.
- Kitchin, R. 2017. Thinking Critically About and Researching Algorithms. *Information, Communication & Society*, 20(1), 14–29.
- Köberer, N. 2014. *Advertorials in Jugendprintmedien. Ein medienethischer Zugang*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Kovach, B., & Rosenstiel, T. 2007. *The Elements of Journalism: What Newspeople should know and the Public should expect*. New York: Three Rivers Press.
- Kraemer, F., Van Overveld, K., & Peterson, M. 2011. Is there an Ethics of Algorithms? *Ethics and Information Technology*, 13(3), 251–260.
- Kroes, P., & Verbeek, P.-P. 2014. *The Moral Status of Technical Artefacts*. Wiesbaden: Springer.
- Kuhlen, R. 1999. *Die Konsequenzen von Informationsassistenten. Was bedeutet informationelle Autonomie?* Frankfurt a.M.: Suhrkamp.
- Kukich, K. 1983. Design and Implementation of a knowledge-based Report Generator. Proceedings of 21st Annual Meeting of the Association for Computational Linguistics, Cambridge, Massachusetts, 15. bis 17. Juni, 145–150. <http://dl.acm.org/citation.cfm?id=981340>.

- Kurzweil, R. 1990. *The Age of Intelligent Machines*. Cambridge (MA): MIT Press.
- Lamnek, S. 2010. *Qualitative Sozialforschung*. Weinheim: Beltz.
- Langer, U. 2011. *Datenjournalismus*. Journalisten Werkstatt. Salzburg-Eugendorf: Medienfachverlag Oberbauer.
- Langford, D. 2000. *Internet Ethics*. New York: St. Martin's Press.
- Latar, N.L. 2015. The Robot Journalist in the Age of Social Physics: The End of Human Journalism? In *The New World of Transitioned Media: Digital Realignment and Industry Transformation*, herausgegeben von Einav, G., 65–80. Wiesbaden: Springer.
- Latour, B. 1987. *Science in Action. How to follow Scientists and Engineers through Society*. Milton Keynes: Open Univ. Press.
- Latzer, M. 2009. Information and Communication Technology Innovations: Radical and disruptive? *New Media & Society*, 11(4), 599–619.
- Latzer, M. 2013. Medienwandel durch Innovation, Ko-Evolution und Komplexität. *Medien & Kommunikationswissenschaft*, 61(2), 235–252.
- Latzer, M., Hollnbuchner, K., Just, N., & Saurwein, F. 2014. The Economics of Algorithmic Selection on the Internet. Working Paper, University of Zurich, Zürich. http://www.mediachange.ch/media/pdf/publications/Economics_of_algorithmic_selection_WP.pdf.
- Latzer, M., Hollnbuchner, K., Just, N., & Saurwein, F. 2016. The Economics of Algorithmic Selection on the Internet. In *Handbook on the Economics of the Internet*, herausgegeben von Bauer, J., & Latzer, M., 395–425. Cheltenham: Edward Elgar.
- Lessig, L. 1999. *Code and Other Laws of Cyberspace*. New York: Basic Books.
- Lewis, S.C. 2011. Journalism Innovation and Participation: An analysis of the Knight News Challenge. *International Journal of Communication*, 5, 1623–1648.
- Lewis, S.C. 2015. Journalism in an Era of Big Data. *Digital Journalism*, 3(3), 321–330.
- Lewis, S.C., & Nikki Usher. 2014. Code, Collaboration, and the Future of Journalism: A Case Study of the Hacks/Hackers Global Network. *Digital Journalism*, 2(3), 383–393.
- Lewis, S.C., & Usher, N. 2013. Open Source and Journalism: Toward new Frameworks for imagining News Innovation. *Media, Culture & Society*, 35(5), 602–619.
- Lewis, S.C., & Westlund, O. 2015a. Actors, Actants, Audiences, and Activities in Cross-Media News Work. *Digital Journalism*, 3(1), 19–37.
- Lewis, S.C., & Westlund, O. 2015b. Big Data and Journalism. Epistemology, Expertise, Economics, and Ethics. *Digital Journalism*, 3(3), 447–466.
- Lin, P. 2015. Why Ethics Matters for Autonomous Cars. In *Autonomes Fahren. Technische, Rechtliche und Gesellschaftliche Aspekte*, herausgegeben von Maurer, M., Gerdes, C.J., Lenz, B., & Winner, H., 69–86. Berlin: Springer.
- Löffelholz, M. 2000. *Theorien des Journalismus. Ein diskursives Handbuch*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Löffelholz, M. 2008. Heterogeneous-Multidimensional-Competing: Theoretical Approaches to Journalism. An Overview. In *Global Journalism Research*, herausgegeben von Löffelholz, M., & Weaver, D., 15–27. Hoboken (NJ): John Wiley & Sons.
- Löffelholz, M., & Rothenberger, L. 2016. *Handbuch Journalismustheorien*. Wiesbaden: Springer.

- Loosen, W. 2015. The Notion of the «Blurring Boundaries». *Digital Journalism*, 3(1), 68–84.
- Loosen, W. 2016. Publikumsbeteiligung im Journalismus. In *Journalismusforschung. Stand und Perspektiven*, herausgegeben von Meier, K., & Neuberger, C., 287–316. Baden-Baden: Nomos.
- Loosen, W., & Scholl, A. 2016. Journalismus im Zeitalter algorithmischer Wirklichkeitskonstruktion. Vortrag auf der 61. Jahrestagung der DGPK, Leipzig.
- Mager, A. 2012. Algorithmic Ideology. *Information, Communication & Society*, 15(5), 269–787.
- Maner, W. 1996. Unique ethical Problems in Information Technology. *Science and Engineering Ethics*, 2(2), 137–154.
- Mason, E.S. 1949. The current Status of the Monopoly Problem in the United States. *Harvard Law Review*, 62(8), 1265–1285.
- Matsuzaki, H. 2011. Die Frage nach der 'Agency' von Technik und die Normenvergessenheit der Techniksoziologie. In *Akteur – Individuum – Subjekt: Fragen zu 'Personalität' und 'Sozialität'*, herausgegeben von Lüdtke, N., & Matsuzaki, H., 301–325. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Matthes, J., & Kohring, M. 2003. Operationalisierung von Vertrauen in Journalismus. *Medien & Kommunikationswissenschaft*, 51(1), 5–23.
- Mayer-Schönberger, V. 2015. Was ist Big Data? Zur Beschleunigung des menschlichen Erkenntnisprozesses. <https://www.bpb.de/apuz/202242/zur-beschleunigung-menschlicher-erkenntnis?p=all>.
- Mayer-Schönberger, V., & Cukier, K. 2013a. *Big Data. Die Revolution, die unser Leben verändern wird*. München: Redline.
- Mayer-Schönberger, V., & Cukier, K. 2013b. *Big Data: A Revolution that will transform how we live, work, and think*. London: John Murray Publishers.
- Mayo, J., & Leshner, G. 2000. Assessing the Credibility of Computer-Assisted Reporting. *Newspaper Research Journal*, 21(4), 68–82.
- McBride, K., & Rosenstiel, T. 2014a. New Guiding Principles for a New Era of Journalism. In *The New Ethics of Journalism*, herausgegeben von McBride, K., & Rosenstiel, T., 1–6. Thousand Oaks (CA): CQ Press.
- McBride, K., & Rosenstiel, T. 2014b. *The New Ethics of Journalism*. Thousand Oaks (CA): CQ Press.
- McCracken, G. 1988. *The Long Interview*. Thousand Oaks (CA): Sage Publications.
- McDonald, D.D. 2010. Natural Language Generation. In *Handbook of Natural Language Processing*, herausgegeben von Indurkha, N., & Damerau, F.J., 121–141. Boca Raton (FL): Chapman and Hall.
- McQuail, D. 2000. *McQuail's Mass Communication Theory*. Thousand Oaks (CA): Sage.
- Meier, K. 2002. Neue Journalistische Formen. In *Internet-Journalismus*, herausgegeben von Meier, K., 21–172. Konstanz: UVK.
- Meier, K. 2003. Qualität im Online-Journalismus. In *Qualität Im Journalismus. Grundlagen, Dimensionen, Praxismodelle*, herausgegeben von Bucher, H.-J., & Altmeyden, K.-D., 247–266. Wiesbaden: Westdeutscher Verlag.
- Meier, K. 2011a. *Journalistik*. Konstanz: UTB.

- Meier, K. 2011b. Journalismusforschung als interaktive Innovationsforschung. Eine Methodologie für Wissenstransfer. In *Methoden der Journalismusforschung*, herausgegeben von Jandura, O., Quandt, T., & Vogelgesang, J., 67–82. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Meier, W.A., & Jarren, O. 2001. Ökonomisierung und Kommerzialisierung von Medien und Mediensystem. Einleitende Bemerkung zu einer (notwendigen) Debatte. *Medien & Kommunikationswissenschaft*, 49(2), 145–158.
- Meuser, M., & Nagel, U. 2009. Das Experteninterview – konzeptionelle Grundlagen und methodische Anlage. In *Methoden der vergleichenden Politik- und Sozialwissenschaft*, herausgegeben von Pickel, S., Pickel, G., Lauth, H.-J., & Jahn, D., 465–479. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Meyen, M. 2009. Das journalistische Feld in Deutschland. Ein theoretischer und empirischer Beitrag zur Journalismusforschung. *Publizistik*, 54(3), 323–345.
- Meyer, P. 2002. *Precision Journalism. A Reporter's Introduction to Social Science Methods*. Lanham: Rowman & Littlefield.
- Mill, J.S. 2006. *Utilitarianism*. Stuttgart: Reclam.
- Minsky, M. 1988. *The Society of Mind*. New York: Simon and Schuster.
- Mitcham, C. 2014. Agency in Humans and Artifacts: A contested Discourse. In *The Moral Status of Technical Artefacts*, herausgegeben von Kroes, P., & Verbeek, P.-P., 11–29. Dordrecht: Springer.
- Mitchelstein, E., & Boczkowski, P.J. 2009. Between Tradition and Change: A Review of recent Research on Online News Production. *Journalism*, 10(5), 562–586.
- Montal, T., & Reich, Z. 2016. I, Robot. You, Journalist. Who is the Author? *Digital Journalism* [online first]. doi: [10.1080/21670811.2016.1209083](https://doi.org/10.1080/21670811.2016.1209083)
- Napoli, P.M. 2014. Automated Media: An Institutional Theory Perspective on algorithmic Media Production and Consumption. *Communication Theory*, 24(3), 340–360.
- Neff, G., Jordan, T., McVeigh-Schultz, J., & Gillespie, T. 2012. Affordances, Technical Agency, and the Politics of Technologies of Cultural Production. *Journal of Broadcasting & Electronic Media*, 56(2), 299–313.
- Neuberger, C. 1996. *Journalismus als Problembearbeitung. Objektivität und Relevanz in der öffentlichen Kommunikation*. Konstanz: UVK.
- Neuberger, C. 2001. *Journalismus im Internet. Theoriekontext und empirische Exploration*. Unveröffentlichte Habilitationsschrift. Eichstätt: Universität Eichstätt-Ingolstadt.
- Neuberger, C. 2002. Vom Unsichtbarwerden des Journalismus im Internet. In *Innovationen Im Journalismus. Forschung für die Praxis*, herausgegeben von Hohlfeld, R., Meier, K., & Neuberger, C., 25–68. Münster: LIT.
- Neuberger, C. 2003. Zeitung Und Internet: Über das Verhältnis zwischen einem alten und einem neuen Medium. In *Online – Die Zukunft der Zeitung?*, herausgegeben von Neuberger, C., & Tonnemacher, J., 16–110. Wiesbaden: Westdeutscher Verlag.
- Neuberger, C. 2004. Journalismus als systembezogene Akteurkonstellation. Grundlagen einer integrativen Journalismustheorie. In *Theorien des Journalismus*, herausgegeben von Löffelholz, M., 287–303. Wiesbaden: VS Verlag für Sozialwissenschaften.

- Neuberger, C. 2007. Interaktivität, Interaktion, Internet. Eine Begriffsanalyse. *Publizistik*, 52(1), 33–50.
- Neuberger, C. 2008. Neue Medien als Herausforderung für die Journalismustheorie: Paradigmenwechsel in der Vermittlung öffentlicher Kommunikation. In *Theorien der Kommunikations- und Medienwissenschaft. Grundlegende Diskussionen, Forschungsfelder und Theorieentwicklungen*, herausgegeben von Winter, C., Hepp, A. & Krotz, F., 251–267. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Neuberger, C. 2009. Internet, Journalismus und Öffentlichkeit. Analyse des Medienumbruchs. In *Journalismus im Internet: Profession – Partizipation – Technisierung*, herausgegeben von Neuberger, C., Nuernbergk, C., & Rischke, M., 19–105. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Neuberger, C. 2012. Bürgerjournalismus als Lösung? Empirische Ergebnisse zu den journalistischen Leistungen von Laienkommunikatoren. In *Medienwandel oder Medienkrise? Folgen für Medienstrukturen und ihre Erforschung*, herausgegeben von Jarren, O., Künzler, M., & Puppis, M., 53–76. Baden-Baden: Nomos.
- Neuberger, C. 2013. Onlinemedien als Institutionen. In *Medien als Institutionen und Organisationen. Institutionalistische Ansätze in der Publizistik- und Kommunikationswissenschaft*, herausgegeben von Künzler, M., Oehmer, F., Puppis, M., & Wassmer, C., 97–116. Baden-Baden: Nomos.
- Neuberger, C. 2014. Die Identität und Qualität des Journalismus im Internet aus der Sicht des Publikums. In *Journalismus und (sein) Publikum*, herausgegeben von Loosen, W., & Dohle, M., 229–251. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Neuberger, C., & Kapern, P. 2013. *Grundlagen des Journalismus*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Neuberger, C., & Nuernbergk, C. 2015. Verdatete Selbstbeschreibung der Gesellschaft – Über den Umgang des Journalismus mit Big Data und Algorithmen. In *Die Gesellschaft der Daten – Über die digitale Transformation der sozialen Ordnung*, herausgegeben von Süßenguth, F., 199–224. Bielefeld: transcript.
- Neverla, I. 2001. Das Netz - Eine Herausforderung für die Kommunikationswissenschaft. In *Kommunikationskulturen zwischen Kontinuität und Wandel. Universelle Netzwerke für die Zivilgesellschaft*, herausgegeben von Maier-Rabler, U., & Latzer, U., 29–46. Konstanz: UVK.
- Neverla, I. 2003. Kommunikationswissenschaft zwischen Komplexität und Kanonisierung. In *Die neue Kommunikationswissenschaft*, herausgegeben von Löffelholz, M., & Quandt, T., 59–68. Wiesbaden: Westdeutscher Verlag.
- Opta. n.d. Covering the World of Football in the Greatest Detail. <http://www.optasports.com/media/836304/final-opta-coverage.pdf>.
- Örnebring, H. 2010. Technology and Journalism-as-Labour: Historical Perspectives. *Journalism*, 11(1), 57–74.
- Örnebring, H., & Ferrer Conill, R. 2016. Outsourcing Newswork. In *The Sage Handbook of Digital Journalism*, herausgegeben von Witschge, T., Anderson, C.W., Domingo, D., & Hermida, A., 207–221. London: Sage.
- Parry, R. 2011. *The Ascent of Media. From Gilgamesh to Google via Gutenberg*. London: Nicholas Brealey Pub.

- Pasquale, F. 2015. *The Black Box Society. The Secret Algorithms that control Money and Information*. Cambridge (MA): Harvard University Press.
- Pavlik, J. 2000. The Impact of Technology on Journalism. *Journalism Studies*, 1(2), 229–237.
- Pavlik, J. 2013. Innovation and the Future of Journalism. *Digital Journalism*, 1(2), 181–193.
- Pérez y Pérez, R., & Sharples, M. 2004. Three computer-based Models of Storytelling: BRUTUS, MINSTREL, and MEXICA. *Knowledge-Based Systems*, 17(1), 15–29.
- Petre, C. 2013. A quantitative Turn in Journalism? <http://towcenter.org/a-quantitative-turn-in-journalism/>.
- Picard, R.G. 1989. *Media Economics: Concepts and Issues*. Newbury Park (CA): Sage.
- Picard, R.G. 2010. A business Perspective on Challenges facing Journalism. In *The Changing Business of Journalism and its Implications for Democracy*, herausgegeben von Levy, D., & Kleis Nielsen, R., 17–24. Oxford: Reuters Institute for the Study of Journalism, University of Oxford.
- Pierce, M.A., & Henry, J.W. 1996. Computer Ethics: The Role of Personal, Informal, and Formal Codes. *Journal of Business Ethics*, 15(4), 425–437.
- Pinch, T.J., & Bijker, W.E. 1984. The Social Construction of Facts and Artefacts: Or how the Sociology of Science and the Sociology of Technology might benefit each other. *Social Studies of Science*, 14(3), 399–441.
- Plesner, U. 2009. An Actor-Network Perspective on Changing Work Practices: Communication Technologies as Actants in Newswork. *Journalism*, 10(5), 604–626.
- Podolny, S. 2015. If an Algorithm wrote this, how would you even know? *New York Times*, 7. März. http://www.nytimes.com/2015/03/08/opinion/sunday/if-an-algorithm-wrote-this-how-would-you-even-know.html?_r=0.
- Portet, F., Reiter, E., Hunter, J., & Sripada, S. 2007. Automatic Generation of textual Summaries from Neonatal Intensive Care Data. *Proceedings of the 11th Conference on Artificial Intelligence in Medicine (AIME 2007)*, 227–236. doi: [10.1007/978-3-540-73599-1_30](https://doi.org/10.1007/978-3-540-73599-1_30)
- Possler, D. & Link, E. 2017. Der Trend Datenjournalismus und seine Bedeutung aus medienethischer Perspektive In *Gesellschaft ohne Diskurs? Digitaler Wandel und Journalismus aus medienethischer Perspektive*, herausgegeben von Stapf, I., Prinzing, M., & Filipović, A., 103–120. Baden-Baden: Nomos.
- Pöttker, H. 1999. Berufsethik für Journalisten? Professionelle Trennungsgrundsätze auf dem Prüfstand? In *Kommunikations- und Medienethik. Interdisziplinäre Perspektiven*, herausgegeben von Holderegger, A., 299–327. Freiburg: Academic Press Fribourg.
- Pöttker, H. 2010. Der Beruf zur Öffentlichkeit. Über Aufgaben, Grundsätze und Perspektiven des Journalismus in der Mediengesellschaft aus der Sicht praktischer Vernunft. *Publizistik*, 55(2), 107–128.
- Powers, M. 2012. In Forms that are familiar and yet-to-be invented: American Journalism and the Discourse of technologically specific Work. *Journal of Communication Inquiry*, 36(1), 24–43.
- Poynter. 2014. L.A. Times Reporter talks about his story-writing 'Quakebot'. <http://www.poynter.org/2014/l-a-times-reporter-talks-about-his-story-writing-quake-bot/243744/>.

- Poynter. 2015. AP Will use Software to write NCAA Game Stories. <http://www.poynter.org/2015/ap-will-use-software-to-write-ncaa-game-stories/324601/>.
- Pratt-Hartmann, I. 2010. Computational Complexity in Natural Language. In *Handbook of Computational Linguistics and Natural Language Processing*, herausgegeben von Clark, A., Fox, C., & Lappin, S., 43–73. Oxford: Wiley-Blackwell.
- Preston, P. 2009. *Making the News: Journalism and News Cultures in Europe*. London & New York: Routledge.
- Primo, A., & Zago, G. 2014. Who and what do Journalism? An Actor-Network Perspective. *Digital Journalism*, 3(1), 38–52.
- Prinzing, M., Rath, M., Schicha, C., & Stapf, I. 2015. *Neuvermessung der Medienethik. Bilanz, Themen und Herausforderungen seit 2000*. Weinheim/Basel: Beltz Juventa.
- Pürer, H. 1992. Ethik in Journalismus und Massenkommunikation. Versuch einer Theorien-Synopse. *Publizistik*, 37(3), 304–321.
- Quandt, T. 2000. Das Ende des Journalismus? Online-Kommunikation als Herausforderung für die Journalismusforschung. In *Theorien des Journalismus. Ein diskursives Handbuch*, herausgegeben von Löffelholz, M., 483–509. Wiesbaden: Westdeutscher Verlag.
- Quandt, T. 2005. *Journalisten im Netz Eine Untersuchung journalistischen Handelns in Online-Redaktionen*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Quandt, T. 2008. Neues Medium, Alter Journalismus? Eine vergleichende Inhaltsanalyse tagesaktueller Print- und Online-Nachrichtenangebote. In *Journalismus Online – Partizipation oder Profession?*, herausgegeben von Quandt, T., & Schweiger, W., 131–155. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Quillian, M.R. 1967. Word Concepts: A Theory and Simulation of some basic semantic Capabilities. *Behavioral Science*, 12, 410–430.
- Rainie, L., & Anderson, J. 2017. Code-Dependent: Pros and Cons of the Algorithm Age. Pew Research Center. <http://www.pewinternet.org/2017/02/08/code-dependent-pros-and-cons-of-the-algorithm-age>.
- Rammert, W. 2003. *Technik in Aktion: verteiltes Handeln in soziotechnischen Konstellationen*. Berlin: Technische Universität, Institut für Soziologie (=Working Papers, 2-2003).
- Rammert, W. 2007. *Technik – Handeln – Wissen. Zu einer pragmatistischen Technik- und Sozialtheorie*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Rammert, W., & Schulz-Schaeffer, I. 2002. Technik und Handeln. Wenn soziales Handeln sich auf menschliches Handeln und technische Abläufe verteilt. In *Können Maschinen handeln? Soziologische Beiträge zum Verhältnis von Mensch und Technik*, herausgegeben von Rammert, W., & Schulz-Schaeffer, I., 11–64. Frankfurt a.M.: Campus.
- Rath, M. 2003a. Medien in Zeiten der Globalisierung – Selbstregulierung zwischen Freiheit und Verantwortung. *Medien Journal*, 27(1), 41–50.
- Rath, M. 2003b. Die medienphilosophische Perspektive: Medien, Wirtschaft, Sinn. In *Medien und Ökonomie. Grundlagen der Medienökonomie*, herausgegeben von Altmeyden, K.-D., & Karmasin, M., 125–139. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Rath, M. 2006. Wahrhaftigkeit des Journalismus: Moralanspruch oder Marktfaktor? In *Medienkritik heute. Grundlagen, Beispiele und Praxisfelder*, herausgegeben von Niesyto, H., Rath, M., & Sowa, H., 117–128. München: Kopaed.

- Redelfs, M. 1996. *Investigative Reporting in den USA. Strukturen eines Journalismus der Machtkontrolle*. Opladen: Westdeutscher Verlag.
- Reich, Z. 2010. Constrained Authors: Bylines and Authorship in News Reporting. *Journalism*, 11(6), 707–725.
- Reinemann, C. 2007. Subjektiv rationale Akteure: Das Potenzial handlungstheoretischer Erklärungen für die Journalismusforschung. In *Journalismustheorie: Next Generation. Soziologische Grundlegung und theoretische Innovation*, herausgegeben von Altmeppen, K.-D., Hanitzsch, T., & Schlüter, C., 47–67. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Reiter, E. 2010. Natural Language Generation. In *The Handbook of Computational Linguistics and Natural Language Processing*, herausgegeben von Clark, A., Fox, C., & Lappin, S., 574–598. Oxford: Wiley-Blackwell.
- Reiter, E., & Dale, R. 2000. *Building Natural Language Generation Systems*. Cambridge: Cambridge University Press.
- Reiter, E., Robertson, R., & Osman, L. 2003. Lessons from a Failure: Generating tailored Smoking Cessation Letters. *Artificial Intelligence*, 144(1-2), 41–58.
- Reiter, E., Sripada, S., Hunter, J., & Yu, J. 2005. Choosing Words in computer-generated Weather Forecast. *Artificial Intelligence*, 167(1-2), 137–169.
- Renger, R. 2006. Zukunft des Content-Zukunft des Journalismus? In *Medien und Ökonomie. Anwendungsfelder der Medienökonomie*, herausgegeben von Altmeppen, K.-D., & Karmasin, M., 145–167. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Resnick, P., & Varian, H.R. 1997. Recommender Systems. *Communications of the ACM*, 40(3), 56–58.
- Rieder, B. 2004. Algorithmische Mediatoren: Software-Agenten als Gegenstand der Medienwissenschaften. *Medien Journal*, 28(1), 36–46.
- Rieder, B. 2005. Networked Control: Search Engines and the Symmetry of Confidence. *International Review of Information Ethics*, 3, 26–32.
- Rip, A. 2002. *Co-Evolution of Science, Technology and Society*. Expert Review, Bundesministerium Bildung und Forschung, Förderinitiative 'Politik, Wissenschaft und Gesellschaft' (Science Policy Studies).
- Rip, A. 2007. Die Verzahnung von technologischen und sozialen Determinismen und die Ambivalenzen von Handlungsträgerschaft im «Constructive Technology Assessment». In *Gesellschaft und die Macht der Technik*, herausgegeben von Dolata, U., & Werle, R., 83–106. Frankfurt: Campus.
- Robin, J., & McKeown, K.R. 1996. Empirically designing and evaluating a new revision-based Model for Summary Generation. *Artificial Intelligence*, 85(1-2), 135–179.
- Ropohl, G. 1991. *Technologische Aufklärung. Beiträge zur Technikphilosophie*. Frankfurt a.M.: Suhrkamp.
- Ross, W.D. 2003. The Right and the Good. In *Deontology*, herausgegeben von Darwall, S., 55–79. Malden: Blackwell.
- Rottwilm, P. 2014. *The Future of journalistic Work: Its changing Nature and Implications*. Oxford: Reuters Institute for the Study of Journalism.

- Rühl, M. 1980. *Journalismus und Gesellschaft. Bestandsaufnahme und Theorieentwurf*. Mainz: v. Hase & Koehler.
- Rühl, M. 1996. Soziale Verantwortung und persönliche Verantwortlichkeit im Journalismus. In *Ethik der Massenmedien*, herausgegeben von Wilke, J., 89–99. Wien: Braumüller.
- Sandvig, C. 2013. The Internet as Infrastructure. In *The Oxford Handbook of Internet Studies*, herausgegeben von Dutton, W.H., 86–106. Oxford (UK): Oxford University Press.
- Saurwein, F., Just, N., & Latzer, M. 2015. Governance of Algorithms: Options and Limitations. *Info*, 17(6), 35–49.
- Schank, R., & Abelson, R. 1977. *Scripts, Plans, Goals, and Understanding*. Hillsdale (NJ): Earlbaum Assoc.
- Schimank, U. 1996. Deutsche Auseinandersetzung mit Luhmann: «Akteurzentrierter Institutionalismus» und gesellschaftliche Differenzierung. In *Theorien gesellschaftlicher Differenzierung*, herausgegeben von Schimank, U., 241–267. Opladen: Leske + Budrich.
- Schimank, U. 2005. *Differenzierung und Integration der modernen Gesellschaft. Beiträge zur akteurzentrierten Differenzierungstheorie*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Schmidt, S.K., & Werle, R. 1998. *Coordinating Technology: Studies in the International Standardization of Telecommunications*. Cambridge (MA): MIT Press.
- Schmitz-Weiss, A., & Domingo, D. 2010. Innovation Processes in Online Newsrooms as Actor-networks and Communities of Practice. *New Media & Society*, 12(7), 1156–1171.
- Scholl, A. 1997. Journalismus als Gegenstand empirischer Forschung: Ein Definitionsvorschlag. *Publizistik*, 42(4), 468–486.
- Scholl, A., & Weischenberg, S. 1998. *Journalismus in der Gesellschaft. Theorie, Methodologie und Empirie*. Opladen: Westdeutscher Verlag.
- Schulten-Jaspers, Y. 2013. *Zukunft der Nachrichtenagenturen. Situation, Entwicklungen, Prognose*. Baden-Baden: Nomos.
- Schulz-Schaeffer, I. 2007. Technik als sozialer Akteur und als soziale Institution: Sozialität von Technik statt Postsozialität. Technische Universität Berlin, TUTS Working Papers 3-2007. <http://nbn-resolving.de/urn:nbn:de:0168-ssoar-12139>.
- Schulz-Schaeffer, I. 2008. Technik als Gegenstand der Soziologie. Technische Universität Berlin, TUTS Working Papers 3-2008. <http://nbn-resolving.de/urn:nbn:de:0168-ssoar-12318>.
- Selgado, S., & Strömbäck, J. 2012. Interpretive Journalism: A Review of Concepts, Operationalizations and Key Findings. *Journalism*, 13(2), 144–161.
- Shirky, C. 2009. A Speculative Post on the Idea of Algorithmic Authority. <http://www.shirky.com/weblog/2009/11/a-speculative-post-on-the-idea-of-algorithmic-authority/>.
- Shirky, C. 2014. Truth without Scarcity, Ethics without Force. In *The New Ethics of Journalism*, herausgegeben von McBride, K., & Rosenstiel, T., 9–24. Thousand Oaks: CQ Press.
- Shoemaker, P.J., & Reese, S.D. 1996. *Mediating the Message: Theories of Influence on Mass Media Content*. White Plains (NY): Longman.
- Shulman, C., Tarleton, N., & Jonsson, H. 2009. Which Consequentialism? Machine Ethics and Moral Divergence. In *AP-CAP 2009: The Fifth Asia-Pacific Computing and Philosophy*

- Conference, 1. bis 2. Oktober, University of Tokyo, Japan, herausgegeben von Reynolds, C., & Cassinelli, A., 23–25. <https://intelligence.org/files/WhichConsequentialism.pdf>.
- Sola Pool, I., & Shulman, I. 1959. Newsmen's Fantasies, Audiences and Newswriting. *Public Opinion Quarterly*, 23(2), 145–158.
- South China Morning Post. 2015. End of the Road for Journalists? Tencent's Robot Reporter 'Dreamwriter' Churns out Perfect 1,000-Word News Story – in 60 Seconds. <http://www.scmp.com/tech/china-tech/article/1857196/end-road-journalists-robot-reporter-dreamwriter-chinas-tencent>.
- Speamann, R. 1977. *Zur Kritik der politischen Utopie*. Stuttgart: Klett.
- Spinello, R.A. 2011. *Cyberethics. Morality and Law in Cyberspace*. Sudbury: Jones & Bartlett.
- Spyridou, L.-P., Matsiola, M., Veglis, A., Kalliris, G., & Dimoulas, C. 2013. Journalism in a State of Flux: Journalists as Agents of Technology Innovation and emerging News Practices. *International Communication Gazette*, 75(1), 76–98.
- Stavelin, E. 2013. *Computational Journalism: When Journalism Meets Programming*. Dissertation, Norwegen: Universität Bergen. <http://bora.uib.no/bitstream/handle/1956/7926/dr-thesis-2013-Eirik-Stavelin.pdf?sequence=1&isAllowed=y>.
- Stewart, J., & Williams, R. 1998. The Coevolution of Society and Multimedia Technology: Issues in predicting the future Innovation and Use of a ubiquitous Technology. *Social Science Computer Review*, 16(3), 268–282.
- Stöber, R. 2008. Innovation und Evolution: Wie erklärt sich medialer und kommunikativer Wandel? In *Theorien der Kommunikations- und Medienwissenschaft. Grundlegende Diskussionen, Forschungsfelder und Theorieentwicklungen*, herausgegeben von Winter, Carsten, Hepp, A., & Krotz, F., 139–156. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Stöber, R. 2013. *Neue Medien. Geschichte: Von Gutenberg bis Apple und Google. Medieninnovation und Evolution*. Bremen: edition lumière.
- Stray, J. 2011. A Computational Journalism Reading List. <http://jonathanstray.com/a-computational-journalism-reading-list>.
- Tavani, H.H. 2011. *Ethics and Technology. Controversies, Questions, and Strategies for Ethical Computing*. New Jersey: Wiley.
- Thomaß, B. 2003. Fünf ethische Prinzipien journalistischer Praxis. In *Kommunikations- und Medienethik*, herausgegeben von Debatin, B., & Funiok, R., 159–168. Konstanz: UVK.
- Thomaß, B. 2016. Ethik des Journalismus. In *Handbuch Journalismustheorien*, herausgegeben von Löffelholz, M., 537–550. Wiesbaden: Springer Fachmedien.
- Thurman, N. 2011. Making 'The Daily Me': Technology, Economics and Habit in the Mainstream Assimilation of Personalized News. *Journalism*, 12(4), 395–415.
- Thurman, N., Dörr, K., & Kunert, J. 2017. When Reporters get Hands-On with Robo-Writing. Professionals consider Automated Journalism's Capabilities and Consequences. *Digital Journalism* [online first]. doi: [10.1080/21670811.2017.1289819](https://doi.org/10.1080/21670811.2017.1289819)
- Thurman, N., Schifferes, S., Fletcher, R., Newman, N., Hunt, S., & Schapals, A.K. 2016. Giving Computers a Nose for News: Exploring the Limits of Story Detection and Verification. *Digital Journalism*, 4(7), 838–848.

- Tuchman, G. 1973. Making News by doing Work. Routinizing the Unexpected. *American Journal of Sociology*, 79(1), 110–131.
- Turi2. 2015. AP Setzt Auf Roboterjournalismus. <http://www.turi2.de/aktuell/ap-setzt-auf-roboterjournalismus/>.
- Van Dalen, A. 2012. The Algorithms behind the Headlines. How machine-written News redefines the Core Skills of Human Journalists. *Journalism Practice*, 6(5-6), 648–658.
- Van der Kaa, H., & Krahmer, E. 2014. Journalist versus News Consumer: The perceived Credibility of Machine Written News. *Vortrag auf dem Symposium Computation + Journalism*, Columbia University, New York, NY. http://compute-cuj.org/cj-2014/cj2014_session4_paper2.pdf.
- von Rimscha, M.B., & Sommer, C. 2015. Fallstudien in der Kommunikationswissenschaft. In *Handbuch nicht standardisierte Methoden in der Kommunikationswissenschaft*, herausgegeben von Averbek-Lietz, S., & Meyen, M., 369–384. Wiesbaden: Springer Fachmedien.
- Vos, T.P., & Heinderyckx, F. 2015. *Gatekeeping in Transition*. New York: Routledge.
- Wachowski, L., & Wachowski, A. (Regisseure). 1999. *Matrix (The Matrix)*. Australien/USA: Warner Bros.
- Ward, S.J. 2006. *The Invention of Journalism Ethics. The Path to Objectivity and Beyond*. Montreal: McGill-Queen's University Press.
- Ward, S.J. 2010. *Global Journalism Ethics*. Montreal: McGill-Queen's University Press.
- Ward, S.J. 2014. The Magical Concept of Transparency. In *Ethics for Digital Journalists: Emerging Best Practices*, herausgegeben von Zion, L., & Craig, D., 45–58. New York: Routledge.
- Weber, M. 1972 [1922]. *Wirtschaft und Gesellschaft. Grundriß der verstehenden Soziologie*. 5., revidierte Aufl. Tübingen: Mohr-Siebeck.
- Weber, W., & Rall, H. 2013. We Are Journalists': Production Practices, Attitudes and a Case Study of the New York Times Newsroom. In *Interaktive Infografiken*, herausgegeben von Weber, W., Burmester, M., & Tille, R., 161–172. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Weeks, L. 2014. Media Law and Copyright Implications of Automated Journalism. *Journal of Intellectual Property and Entertainment Law*, 4(1), 67–94.
- Weinacht, S., & Spiller, R. 2014. Datenjournalismus in Deutschland. Eine explorative Untersuchung zu Rollenbildern von Datenjournalisten. *Publizistik*, 59(4), 411–433.
- Weischenberg, S. 1985. Die Unberechenbarkeit des Gatekeepers. Zur Zukunft professioneller Informationsvermittlung im Prozess technisch-ökonomischen Wandels. *Rundfunk und Fernsehen*, 33(2), 187–201.
- Weischenberg, S. 1992. *Journalistik. Band 1: Mediensysteme, Medienethik, Medieninstitutionen*. Opladen: Westdeutscher Verlag.
- Weischenberg, S. 1995. *Journalistik. Band 2: Medientechnik, Medienfunktionen, Medienakteure*. Opladen: Westdeutscher Verlag.
- Weischenberg, S. 2001. *Nachrichten-Journalismus. Anleitungen und Qualitäts-Standards für die Medienpraxis*. Wiesbaden: Westdeutscher Verlag.

- Weischenberg, S. 2006a. Medienqualitäten: Zur Einführung in den kommunikationswissenschaftlichen Diskurs über Maßstäbe und Methoden zur Bewertung öffentlicher Kommunikation. In *Medien-Qualitäten. Öffentliche Kommunikation zwischen ökonomischem Kalkül und Sozialverantwortung*, herausgegeben von Weischenberg, S., Loosen, W., & Beuthner, M., 9–36. Konstanz: UVK.
- Weischenberg, S. 2006b. Qualitätssicherung—Qualitätsstandards für Medienprodukte. In *Handbuch Medienmanagement*, herausgegeben von Scholz, C., 663–685. Berlin: Springer.
- Weischenberg, S. 2014. *Max Weber und die Vermessung der Medienwelt. Empirie und Ethik des Journalismus – eine Spurenlese*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Weischenberg, S., & Hienzsch, U. 1994. Die Entwicklung der Medientechnik. In *Die Wirklichkeit der Medien. Eine Einführung in die Kommunikationswissenschaft*, herausgegeben von Merten, K., Schmidt, S.J., & Weischenberg, S., 455–480. Opladen: Westdeutscher Verlag.
- Weischenberg, S., Malik, M., & Scholl, A. 2006a. *Die Souffleure der Mediengesellschaft: Report über die Journalisten in Deutschland*. Konstanz: UVK.
- Weischenberg, S., Malik, M., & Scholl, A. 2006b. Journalismus in Deutschland 2005. Zentrale Befunde der aktuellen Repräsentativbefragung deutscher Journalisten. *Media Perspektiven*, 7, 346–361.
- Weischenberg, S., Malik, M., & Scholl, A. 2012. Journalism in Germany in the 21st Century. In *The Global Journalist in the 21st Century*, herausgegeben von Weaver, D.H., & Willnat, L., 205–219. New York: Routledge.
- Weiss, R.S. 1994. *Learning from Strangers*. New York: The Free Press.
- Werle, R. 2002. Technik als Akteurfiktion. In *Können Maschinen handeln? Soziologische Beiträge zum Verhältnis von Mensch und Technik*, herausgegeben von Rammert, W., & Schulz-Schaeffer, I., 119–139. Frankfurt a.M.: Campus.
- Werle, R. 2005. Institutionelle Analyse technischer Innovation. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 57, 308–332.
- Werle, R. 2007. Zur Interdependenz von Innovationen. In *Innovationsforschung*, herausgegeben von Hof, H., & Wengenroth, U., 31–40. Hamburg: LIT.
- Weßler, H. 2002. Journalismus und Kommunikationswissenschaft. Eine Einleitung. In *Journalismus – Medien – Öffentlichkeit. Eine Einführung*, herausgegeben von Jarren, O., & Weßler, H., 17–38. Wiesbaden: Westdeutscher Verlag.
- Whitby, B. 1996. *Reflections on Artificial Intelligence: The legal, moral, and ethical Dimensions*. Exeter (UK): Intellect.
- Whitney, D.C, Sumpter, R.S., & McQuail, D. 2004. News Media Production: Individuals, Organizations, and Institutions. In *The SAGE Handbook of Media Studies*, herausgegeben von Downing, J.D.H., McQuail, D., Schlesinger, P., & Wartella, E.A., 393–409. Thousand Oaks (CA) & London: Sage.
- Wiesenberg, M., Zerfass, A., & Moreno, A. 2017. Big Data and Automation in Strategic Communication. *International Journal of Strategic Communication*, [online first]. doi: [10.1080/1553118X.2017.1285770](https://doi.org/10.1080/1553118X.2017.1285770)
- Willnat, L., Weaver, D.H., & Choi, J. 2013. The Global Journalist in the Twenty-First Century: A Cross-National Study of Journalistic Competencies. *Journalism Practice*, 7(2), 163–183.

- Winner, L. 1986. *The Whale and the Reactor*. Chicago: University of Chicago Press.
- Wolf, C. 2014. *Mobiler Journalismus. Angebote, Produktionsroutinen und redaktionelle Strategien Deutscher Print- und Rundfunkredaktionen*. Baden-Baden: Nomos.
- Wunden, W. 2003. Die «Publizistischen Grundsätze» des Deutschen Presserats aus medienethischer Sicht. In *Kommunikations- und Medienethik*, herausgegeben von Debatin, B., & Funiok, R., 169–182. Konstanz: UVK.
- Wyss, V., & Keel, G. 2010. Journalismusforschung. In *Einführung in die Publizistikwissenschaft*, herausgegeben von Bonfadelli, H., Jarren, O., & Siegert, G., 337–378. Bern: Haupt.
- Yarnall, L., Johnson, J.T., Rinne, L., & Ranney, M.A. 2008. How Post-Secondary Journalism Educators teach advanced CAR Data Analysis Skills in the Digital Age. *Journalism & Mass Communication Educator*, 63(2), 146–164.
- Yin, R.K. 2009. *Case Study Research. Design and Methods*. Los Angeles (CA): Sage Publications.
- Young, M.L., & Hermida, A. 2015. From Mr. and Mrs. Outlier to central Tendencies. Computational Journalism and Crime Reporting at the Los Angeles Times. *Digital Journalism*, 3(3), 381–397.
- Yu, J., Reiter, E., Hunter, J., & Mellish, C. 2007. Choosing the Content of textual Summaries of large Time-Series Data Sets. *Natural Language Engineering*, 13(01), 25–49.
- Zion, L., & Craig, D. 2014. *Ethics for Digital Journalists. Emerging Best Practices*. New York: Routledge.

Anhang

A. Interviews und Transkripte

Im Anhang dieser Arbeit werden ausschließlich diejenigen Interviews und Transkriptionen chronologisch aufgeführt, die im Rahmen der Dissertation noch nicht verwertet wurden. Sie wurden zu unterschiedlichen Zeitpunkten des Dissertationsprojekts erhoben, um einen Ein- und Überblick über das Themengebiet, spezifische Problemdimensionen und das Potential der Technologie zu erarbeiten. Auf diese ergänzenden Interviews wird in der Synopse verwiesen. Die Transkriptionen der Interviews und die Audiodateien der Artikel II und III liegen vollständig vor und können bei Bedarf beim Autor angefragt werden.

A.1 Manfred Stede, Prof. für Angewandte Computerlinguistik, Universität Potsdam

Interviewer: Konstantin Dörr (KD)

Interviewpartner: Manfred Stede (MS), Prof. für Angewandte Computerlinguistik, Univ. Potsdam

Form des Interviews: schriftlich

Datum: 05.01.2015

Sprache: Deutsch

Algorithmischer Journalismus – Grenzen und Potential der Natural Language Generation

Sehr geehrter Herr Prof. Stede,

ich darf mich an dieser Stelle noch einmal herzlich für Ihre Unterstützung bedanken. Wie ich bereits in unserer Mail-Korrespondenz skizziert habe, nimmt die Relevanz algorithmischer Textgenerierung für die strategische und wirtschaftliche Ausrichtung vieler Medienorganisationen und unseren täglichen Medienkonsum zu. Anbieter (z.B. aexea) und Kunden (z.B. Sport-Informationen-Dienst (SID)) drängen auf das journalistische Hoheitsgebiet der Inhaltserstellung und experimentieren bereits mit journalistischen Angeboten. «Algorithmischer Journalismus» ist dabei in seinem Kern computerlinguistische Textgenerierung (NLG).

Die Natural Language Generation wird im Journalismus verstärkt in den Bereichen der Sportberichterstattung (Live-Ticker), Wetter- und Finanzberichterstattung eingesetzt, weil hier elektronische Daten in strukturierter Form für eingrenzbare Domänen (Kontext) vorliegen.

Um diese Entwicklung erfassen und einschätzen zu können, ist es wichtig, sich mit der technischen Funktionsweise und dem Potential der Textgenerierung in der Computerlinguistik und deren Auswirkung auf den Journalismus auseinanderzusetzen.

Dabei geht es sowohl um die technische Ebene der Textgenerierung (Potential und Grenzen) in der Computerlinguistik als auch um Ihre persönliche Experteneinschätzung zu den Grenzen automatisierter Textgenerierung – besonders im Hinblick auf den Einsatz im Journalismus. Für die Beantwortung der Fragen gibt es keine Zeichen- oder Wortbegrenzung. Frage(n): Wo sehen Sie die technischen Grenzen, aber auch das Potential der Textgenerierung allgemein und speziell im Hinblick auf eine journalistische Verwertbarkeit bzw. einen journalistischen Einsatz?

Dabei sind folgende Teilfragen von Interesse:

1. *Ambiguitätsproblematik und Weltwissen in der Textgenerierung: Welche Grenzen und welches Potential der NLG ergeben sich Ihrer Meinung nach dadurch für eine journalistische Verwertbarkeit? Welche zentralen Problemfelder der NLG schätzen Sie darüber hinaus in diesem Themenfeld als relevant ein?*

MS: Ich würde Ambiguität und Weltwissen als «Problemaspekte» auseinanderhalten. Weltwissen in der Gestalt von «Fakten» ist inzwischen in strukturierter Form in beträchtlichem Maß vorhanden (große Ontologien, WikiData etc.) und kann vergleichsweise einfach in Generierungssysteme eingebunden werden. Zu beachten ist die Einschränkung auf «Fakten» – andere Wissensformen (episodisches Wissen, Erfahrungen, ...) sind bisher nur ansatzweise formalisiert und nicht in größerem Umfang verfügbar.

Ambiguität ist eine ganz normale Eigenschaft natürlicher Sprache; sie ist nicht tragisch, sondern sogar wertvoll. Relativ selten sind in «natürlichen» Texten Fälle, wo man etwas auf sehr unterschiedliche Weise verstehen kann und davon dann auch einiges abhängt. Textgeneratoren werden gelegentlich Ambiguitäten in Texten erzeugen, in der Regel ohne das zu bemerken. In den meisten Fällen werden auch die Leser das nicht bemerken, weil die Inhaltsdomäne oder der Kontext die intendierte Lesart nahelegen. Komplikationen mögen gelegentlich auftreten, aber Ambiguität scheint mir insgesamt nicht das dringendste Problem der Generierung.

2. *Journalistische Darstellungsformen – Nachricht vs. Reportage: Wie lang und komplex können Texte (theoretisch) sinnhaft generiert werden, wenn die Datenlage ausreichend ist? Welche Bedingungen müssen erfüllt werden, bzw. wo zeigen sich technische und computerlinguistische Grenzen der journalistischen Darstell- und Verwertbarkeit? Ist es möglich bzw. absehbar, dass eine Software z.B. eine ausführliche Reportage (Zitate, bildliche Beschreibungen, Schlussfolgerungen etc.) über Obdachlose in Berlin generieren kann (Stichwort: Domänen bzw. Kontexte)? Wenn nicht, was sind die Gründe?*

MS: Textlänge ist nicht entscheidend, ein Sportbericht über ein Fußballspiel kann nach Belieben sehr kompakt oder sehr ausführlich generiert werden. Solange es nur um Fakten geht, spielt die Textlänge eigentlich keine Rolle. Ganz wichtig hingegen der Unterschied

zwischen Nachricht und Reportage. Ich glaube nicht, dass die heute populären Demonstratoren für Nachrichtentexte (die ja alle auf strukturierten Daten beruhen) in naher Zukunft in ähnlicher Weise auch (wirklich gut lesbare) Reportagen hervorbringen können. Für eine Reportage muss ich als Autor nicht nur Fakten aufzählen, sondern muss Fakten bewerten, sie auf eine Art und Weise organisieren, dass es meiner Gesamtintention entspricht. Und, wie Sie richtig sagen, zu den «Fakten» vieles anderes hinzutun, dessen Bedeutung wiederum erst im Kontext und aufgrund der Autoren-Absicht entsteht. Letztlich liegt einer Reportage auf Seiten des Autors eine «Haltung» zugrunde. Textgeneratoren haben einstweilen keine.

3. *Ist und, wenn ja, bis zu welchem Ausmaß kann Textgenerierung «kreativ» bzw. unerwartbar sein (Satzkonstruktion, Wortwahl etc.)? Wo sind die Grenzen der textlichen Kreativität (Stichwort: lernende Algorithmen)? Ist «Kreativität» nur Teil des Programmcodes und muss je nach Textart und -form wiederholt in Feedbackschleifen bis zum gewünschten Textergebnis angepasst werden oder müssen Journalisten fürchten, in absehbarer Zeit von einer Software ersetzt zu werden?*

MS: Kreativität auf der rein sprachlichen Ebene ist möglich, wenn auch gefährlich – der Generator kann nur schwer abschätzen, was denn noch in der intendierten Weise «verständlich» ist und was nicht. Durch maschinelles Lernen bzw. Statistik auf der Grundlage großer Textkorpora lässt sich das teilweise kompensieren.

Kreativität auf der inhaltlichen Ebene war vor langer Zeit schon mal Forschungsgegenstand unter dem Schlagwort «story generation». Entsprechende Systeme haben ganz interessante Dinge produziert, aber nur in sehr eng begrenzten Domänen, die eben komplett manuell modelliert sein mussten.

Ich glaube nicht, dass die Kreativitätsfrage so eng mit der Ersetzbarkeit des Journalisten zusammenhängt. Es ist am Ende wieder die Frage der Textsorte. Nachrichten im weitesten Sinne (aus Daten) bedürfen halt keiner großen Kreativität. Kommentare, Reportagen, Feuilleton sind ganz andere Texte, wo es auf die oben genannte «Haltung» ankommt. Kein Einsatzfeld für Maschinen, denke ich.

4. *Ist es möglich, Texte mit unterschiedlichen Sprachstilen und Tonalitäten (z.B. Boulevardstil vs. Nachrichtenstil) zu generieren? Welche Voraussetzungen sind dazu notwendig (Logiken, Lexika & Grammatik im Programmcodes etc.)? Als wie komplex ist dieses Verfahren aus technischer und linguistischer Sicht einzuschätzen? Wo sind die Grenzen dieser sprachlichen Darstellung?*

MS: Dies wurde in den 1990er Jahren bearbeitet. Googeln Sie «Pauline» (Generator aus der Dissertation von E. Hovy). Ich habe selbst auch dazu gearbeitet und ein wenig publiziert. Aufbauend auf meiner Diss. hat Philipp Edmonds mit seinem System «I-Saurus» (auch

googlebar) verschiedene Stil-Ebenen generiert. Diese Arbeiten sind dann in den Hintergrund geraten, weil der Statistik-Boom die NLG übernommen hat. Könnte aber sehr gut sein, dass die nun wieder aktuell werden.

5. *Wie realistisch ist die Aussage eines Inhaltsanbieters aus technischer und linguistischer Sicht, dass journalistische Texte von einem Textgenerierungsprogramm in unterschiedlichen Sprachen (Deutsch, Englisch, Französisch etc.) angeboten werden können? Was ist dazu notwendig?*

MS: Prinzipiell nicht schwer, wenn es wieder um Nachrichten bzw. um Sachtexte geht. Ich habe in meiner Diss. (1990er) mehrsprachige Bedienungsanleitungen für Autos generiert. Es ist eine Frage der Trennung von Weltwissen, Domänenwissen, sprachlichem Wissen, das muss man sorgfältig separat modellieren.

6. *Eingrenzbare Domänen (Themenkontexte) sind ein Möglichkeitsrahmen für journalistische Darstellungsvarianten (z.B. Ticker für Fußball, 3. Liga): Wo sehen Sie darüber hinaus ein Potential der Textgenerierung für einen Einsatz im Journalismus?*

MS: Na ja, das ist die 1-Million-Euro-Frage, die müssen die jungen Firmen beantworten. :)

A.2 Jens Wagner, Leiter Technik und Projekte, Sport-Informations-Dienst (SID)

Interviewer: Konstantin Dörr (KD)

Interviewpartner: Jens Wagner (JW), Leiter Technik & Projekte, Sport-Informations-Dienst (SID)

Form des Interviews: schriftlich

Datum: 06.01.2015

Sprache: Deutsch

Interview: Algorithmischer Journalismus – Sport-Informations-Dienst

Sehr geehrter Herr Wagner,

ich darf mich an dieser Stelle noch einmal herzlich für Ihre Unterstützung bedanken. Wie ich bereits in unserer Mail-Korrespondenz skizziert habe, nimmt die Relevanz automatisierter Textgenerierung für die strategische und wirtschaftliche Ausrichtung vieler Medienorganisationen und unseren täglichen Medienkonsum zu. Auch der SID ist Vorreiter auf diesem Gebiet. Gerade die Sportberichterstattung bietet sich aufgrund der strukturierten Datenlage für diese Technologie an.

Für meine Dissertation führe ich Marktanalysen bei internationalen Anbietern dieser Software-Lösungen und Kunden aus dem Journalismus durch. Dabei geht es um die Frage, wie diese Form des Journalismus in das jeweilige Geschäftsmodell integriert ist und welches Potential sich daraus ergibt. Daneben interessiert mich, welche Gründe im Unternehmen dazu führen, dass man sich für ein Produkt entscheidet, welche Erfahrungen bereits mit der Software gemacht wurden und welche Produkte darüber hinaus existieren oder geplant sind.

Im Anhang finden Sie den Fragebogen, dessen Beantwortung keine Begrenzung in Zeichen- oder Wortzahl vorsieht. Sie können direkt in dieses Dokument schreiben und es retour an mich an die Adresse k.doerr@ipmz.uzh.ch senden (bitte das Speichern des Dokuments nicht vergessen).

Falls Sie ausdrücklich wünschen, dass eine Information nicht wissenschaftlich verwendet werden soll, bitte ich Sie das im Text zu kennzeichnen. Dies wird selbstverständlich berücksichtigt.

Ich bin über jede Art von Information dankbar, um das Bild und das Potential des Algorithmischen Journalismus für die Journalismusforschung einschätzen zu können und weiß Ihre Mühe sehr zu schätzen.

Herzlichen Dank für Ihre Unterstützung! Bei Fragen melden Sie sich bitte jederzeit.

Mit freundlichen Grüßen

Konstantin Dörr

Fragebogen Sport-Informations-Dienst (SID)

1. *Ist bereits oder wird in naher Zukunft die Textgenerierung (Terminankündigung) in Zusammenarbeit mit aexea den Status eines Pilotprojekts verlassen und in das journalistische Angebot des SID integriert und so kommerziell nutzbar? Wie ist der Stand der Dinge?*

JW: Das Ziel des Pilotprojektes ist es, zu prüfen ob:

- a) die Qualität der generierten Texte unseren Ansprüchen genügt,
- b) diese Qualität den Ansprüchen unserer Kunden genügt,
- c) Kunden willens sind, diese Texte zu verwenden und zu bezahlen.

Insofern hängt die Überführung des Projekts zu einem Produkt von diversen Faktoren ab. Da wir uns derzeit noch in einer Phase befinden, die vor der Kundenangebotsphase liegt, ist über die Zukunft derzeit noch keine Aussage zu treffen.

2. *Gibt es aktuell weitere Projekte und Produkte des SID oder sind weitere Produkte geplant, die sich mit automatisierter Textgenerierung befassen? Wenn ja, welche? Bitte beschreiben Sie den «groben» Funktionsumfang und das geplante Einsatzgebiet, z.B. Live-Ticker, regionale Auswahl an Nachrichten, Eingrenzung auf spezifisch ausgewählte Sportarten und Ligen.*

JW: Das ist derzeit nicht geplant und hängt vom Ausgang des Pilotprojektes ab.

3. *Bitte beschreiben Sie Ihre Erfahrung: Können Produkte der Textgenerierung dem Qualitätsanspruch des SID zur journalistischen Professionalität gerecht werden?*

JW: Ganz klar: Nein! Zumindest, wenn bei dieser Frage die Betonung auf «journalistisch» liegt und damit die journalistischen Qualitäten bezüglich Bewertungen, Verknüpfungen und Kreativität gemeint sind. Diese automatisch generierten Texte sind keine journalistischen Texte, sondern die Versprachlichung von Datenbankinhalten. Entsprechend werden sie in dieser Form nie unserem journalistischen Qualitätsanspruch gerecht werden können.

Im Rahmen der journalistischen Professionalität, Korrektheit und Stringenz sieht das anders aus: Unser Ziel ist es, mit den generierten Texten hier unsere Ansprüche bezüglich Inhalt und Form zu erfüllen. Dabei natürlich auch Korrektheit und Stringenz.

4. *Bitte beschreiben Sie Ihre Erfahrung: Auf welchen Ebenen liefern Produkte der Textgenerierung dem Kunden einen Mehrwert im Vergleich zum traditionellen Angebot des SID (Vertrauensbildung)?*

JW: Wir werden im Pilotprojekt Texte anbieten, die derzeit nicht von uns redaktionell erstellt werden. Heute gibt es die Terminankündigungen ausschließlich als Datenlisten.

Der Mehrwert liegt somit in der Form: Es können Termine nun auch in Textform von uns bezogen werden. Ein weiterer Mehrwert liegt in der möglichen Individualisierung für einen Kunden. Alle Nachrichtenagenturen liefern traditionell EINEN Text an VIELE Kunden.

Solange traditionelle Kunden diese Kunden in ihren lokalen und regionalen Medien genutzt haben, waren Doppelungen hier nie ein Problem. Durch die Nutzung im Internet und das Gewicht des Google-Rankings ist allerdings *duplicate content* ein immer größer werdendes Problem. Dieses Problem werden wir somit zumindest zum Teil angehen können.

5. *Welche Kunden des SID haben Produkte auf Grundlage automatisierter Texterkennung bereits abonniert? Falls NDAs: Wie ist die Verteilung von Online-Kunden, Print-Kunden, Print/Online-Kunden mit journalistischer Kernleistung in Prozent je nach Produkt?*

JW: Noch bieten wir das Produkt nicht an.

6. *Aus Kundensicht: Muss für die Nutzung der Textgenerierungsprodukte eine spezielle Lizenz kostenpflichtig erworben werden, um die Leistung zu nutzen, oder ist die Leistung in das allgemeine Abo (SID-Text) inkludiert? Wie ist die Preisstruktur des Angebots?*

JW: Es wird ein gesondertes Produkt sein und auf alle Fälle sowohl separat zu buchen als auch zu bezahlen sein. Die Preisstruktur ist derzeit noch nicht festgelegt. Wir diskutieren derzeit verschiedene Modelle: Flatrates und Per-Story-Rates.

7. *In welchen Sprachen werden Produkte der Textgenerierung für Kunden angeboten oder in welchen ist es geplant?*

JW: Da wir ausschließlich deutsche Dienste anbieten, werden wir mit deutschen Texten starten. Unsere Mutter AFP ist international aufgestellt und schaut interessiert auf unsere Erfahrungen. Sobald das Projekt läuft werden wir sicher auch Tests in anderen Sprachen (EN, FR, ES, ...) machen.

8. *Für welchen Markt werden Produkte der automatisierten Textgenerierung im SID hauptsächlich entwickelt? Welche Länder und welche Sprachräume stehen im Fokus?*

JW: Siehe oben.

9. *Gibt es eine konkrete Nachfrage auf Kundenseite nach diesen Produkten? Wenn ja, was wird konkret gewünscht und nachgefragt (z.B. automatisierte Kurznachrichten, Live-Ticker etc.)?*

JW: Die konkrete Nachfrage bezieht sich üblicherweise immer auf individualisierten Content. Nie explizit auf generierten Content. Und im Liveticker-Bereich produzieren wir schon seit langer Zeit selbst automatisierte Feeds z.B. für Torschützen und Halbzeit-/Endstände.

10. *Bitte schätzen Sie ein: Welchen Stellenwert hat die automatisierte Textgenerierung innerhalb des Geschäftsmodells des SID aktuell und in Zukunft? Komplementär und auf gleicher Ebene wie das traditionelle Angebot des SID-Texts?*

JW: Aktuell: Null Prozent. Ausgehend von einer sinnvollen Umsetzung und Akzeptanz am Markt, entsprechender Erweiterung auf Ergebnismeldungen o.Ä. kann ich mir vorstellen, dass wir einen einstelligen Umsatz-Prozentsatz mit generierten Texten erwirtschaften können. Wobei als Grundlage immer entsprechend aufbereitete Datenbanken bereitstehen müssen und wir im Projekt schon merken, dass deren Pflege sich deutlich gegenüber unserer heutigen DB-Pflege unterscheidet.

11. *Bitte schätzen Sie ein, welche Auswirkungen der Ausbau und die Integration der automatisierten Textgenerierung auf die Personalstruktur des SID haben könnten oder bereits haben (z.B. Rationalisierungen des journalistischen Personals).*

JW: Wir werden keinerlei journalistische Produktion durch generierte Texte ersetzen. Weder können, noch wollen. Generierte Texte werden immer Add-Ons bleiben. Somit sehe ich zusätzlichen Personalbedarf im Bereich Datenpflege.

12. *Wie viele Mitarbeiter sind am Ausbau und an der Integration mit Projekten der automatisierten Texterstellung beim SID involviert? Wie ist der Anteil der Mitarbeitenden in Teil-/Vollzeit?*

JW: Derzeit sind 3 Vollzeitmitarbeiter und diverse Freie mit dem Thema befasst.

13. *Haben Sie außerhalb Ihrer Kooperation mit aexea weitere Partner in diesem Geschäftsfeld, z.B. text-on, Automated Insights, Narrative Science, Linguastat etc.? Wenn ja, welche?*

JW: Derzeit nicht.

14. *Inwiefern hat die AFP als Muttergesellschaft Anteil und Mitspracherecht an der strategischen Ausrichtung des SID im Bereich der Textgenerierung? Hat sie Verwertungsrechte an Produkten?*

JW: AFP ist interessiert an den Ergebnissen des Projekts. Wobei natürlich ein Fokus die Multilingualität (siehe oben) ist. Natürlich kann unsere Muttergesellschaft unsere Produkte verwerten. Da wir aber den deutschen Markt abdecken, ist für deutschsprachige Inhalte weltweit der Markt vernachlässigbar klein.

15. *Die Grundlage jeder Art von Textgenerierung sind strukturierte Datensätze: Werden die Daten für eventuelle Produkte durch den SID erhoben und aufbereitet oder werden sie von einem externen Anbieter eingekauft? Wenn zugekauft, wie heißen die Datenlieferanten und, falls möglich, in welchem Preisrahmen bewegen wir uns?*

JW: Aus Zuverlässigkeitsaspekten verwenden wir ausschließlich eigene Datenbanken.

16. *Für welche Bereiche der Sportberichterstattung hat der SID Zugriff auf elektronische Daten/Datenbanken (Fußball, Tennis, Basketball etc.), die einen Einsatz für automatisierte Textangebote ermöglichen könnten?*

JW: Im Prinzip für alle olympischen Sportarten. Allerdings muss jeweils geprüft werden, ob wir die nötige Datentiefe vorhalten. Dann ist von Fall zu Fall zu entscheiden, ob es wirtschaftlich vertretbar ist, diese zu erweitern, um daraus neue Produkte erstellen zu können.

17. *Werden die Datenbanken auf Objektivität, Glaubwürdigkeit und Verlässlichkeit geprüft, bevor ein journalistisches Produkt bei aexea oder einem anderen Partner entsteht? Gibt es so etwas wie einen Code of Conduct, der Normen, Werte und Umgang mit Daten zur Erstellung journalistischer Produkte regelt?*

JW: Der Vorteil eines kleinen Unternehmens wie des SID ist, dass wir mit einem überschaubaren Team arbeiten. Unser Datenteam arbeitet immer unter der Prämisse der Korrektheit und Geschwindigkeit. Diese beiden Faktoren machen seit jeher die Glaubwürdigkeit des SID und seiner Produkte aus. Dementsprechend werden wir nie ein automatisch generiertes Produkt auf fremden Daten aufbauen – so uns deren Verlässlichkeit nicht 100 Prozent überzeugt.

18. *Werden Produkte der Textgenerierung des SID mit einer Signalisierung/einem Branding versehen, z.B. «Diese Terminankündigung wurde von der Software xy erstellt», oder kann der Kunde entscheiden, wie er mit den Inhalten umgeht und sie in sein Angebot einbindet?*

JW: Wir planen derzeit einen entsprechenden Disclaimer. Dieser wird aber vom Kunden nicht ausgespielt werden müssen.

A.3 Carsten Erdmann, Chefredakteur, Berliner Morgenpost

Interviewer: Konstantin Dörr (KD)

Interviewpartner: Carsten Erdmann (CE), Chefredakteur, Berliner Morgenpost

Interviewform: Face to Face, Redaktionsräume der Berliner Morgenpost in Berlin

Datum: 24.04.2015

Dauer: 00:53:09 Minuten

Sprache: Deutsch

[Smalltalk: Einleitung, Schweizer Medienlandschaft, NZZ, Feinstaubmonitor, Struktur des Interviews] von #00:00:00# bis #00:01:18#

KD: Es würde mich grundsätzlich interessieren, ob in naher Zukunft neben dem Feinstaubmonitor bei der Berliner Morgenpost andere journalistische geplant sind, die auf automatisierter Textgenerierung basieren. #00:01:41#

CE: Das Projekt «Feinstaubmonitor» ist bei unserem Innovations-Team, aus unserem Interaktiv-Team entstanden. Wir haben ein Team, in dem Datenjournalisten, Web-Developer und Journalisten im Newsroom zusammenarbeiten. Es ist ein kleines Team, aber es ist sehr effizient. Wir waren ein bisschen von der LA Times inspiriert, die den Mut bei dem Erdbeben gehabt hat ... Dort haben wir das gesehen. Wir haben uns gefragt: Gibt es bei uns Daten, mit denen wir automatisiert arbeiten können? Dann haben wir uns gefragt: Welches Thema eignet sich in Berlin? Was ist denn so ein Berliner Thema? Wetter, ja gut ... das kann man immer machen. Aber Feinstaub war dann genau so ein Thema. Ja, es gibt eine Belastung, und da gibt es Daten, wie könnten wir das scrapen? Und dann haben wir uns für das Thema Feinstaub entschieden. Es war aber eher auch für uns ein Test. Funktioniert das eigentlich? Was kommt dabei raus? Wir haben auch gesehen, dass das auch schon Kapazitäten kostet. Wir sind eine Regionalzeitung und haben ein kleines Daten-Team – ich habe hier einen Newsroom und selbst das ist für eine Regionalzeitung ungewöhnlich, dass da Developer sitzen, die mit den Journalisten zusammenarbeiten. Normalerweise sitzen die ja eher auf der Produktseite. Die haben jetzt dieses Projekt gemacht. Aber das sprengt künftige Sachen, wenn man das weiterdenken will. Wir haben also den Feinstaubmonitor gelauncht. Da gab es auch relativ viel Presse, weil es offenbar noch nicht so viele machen. Und die Presse galt von «innovativ» bis «jetzt kommen die Roboter und vernichten alle Arbeitsplätze», also sehr, sehr stark polarisiert. Natürlich haben wir uns dann zwischenzeitlich wieder mit vielen anderen Dingen beschäftigt. Aber wir haben uns schließlich überlegt, in welchen Bereichen – weil das Projekt ganz gut gelaufen ist und auch keine großen Fehler gezeigt hat – könnten wir neue Projekte machen? Dann haben wir zwei oder drei Felder eingegrenzt, in denen man arbeiten könnte. Das eine ist der Bereich Wirtschaft. Das andere ist der Bereich Service/Verkehr/Wetter. Und das Dritte ist der Bereich Sport. Wir haben uns dann verschiedene Anbieter angesehen. Mit text-on arbeiten wir aktuell an

einem Wirtschaftsprodukt für Berlin. Und wir arbeiten auch – und das ist wirklich in der Frühphase – an einem Amateursportprodukt. Dabei haben wir natürlich auch gesehen, dass es nicht ganz einfach ist. Das wichtigste ist ja immer, dass wir diese Daten irgendwie in einer verlässlichen Form bekommen. Das kann ich in der Wirtschaft noch einkaufen. Aber im Sportbereich – wenn ich es dann wirklich für eine Regionalmarke spannend haben will – gibt es das noch nicht. Jedenfalls nicht so, wie wir uns das vorstellen, dass wir wirklich größere Textstücke daraus machen können. Und wenn es diese gibt, stellt sich die Frage, wie lange uns z.B. ein Verband wie der DFB dies kostenlos zur Verfügung stellt? Weil es auch dort wieder Bestrebungen gibt, selbst etwas zu machen oder zu vermarkten. Also der Open-Data-Gedanke ist dort noch nicht so ausgeprägt. #0:6:15.5#

[Smalltalk: Gliederung des Interviews] von #0:6:15.5# bis #0:6:18.6#

[Off the record: Fußballberichterstattung] von #0:6:18.6# bis #0:9:0.9.#

KD: Sie planen weitere Produkte mit text-on? #0:9:3.7#

CE: Ja, mit text-on fahren wir jetzt mal einen 3-Monats-Test. Wir sehen uns an, wie zuverlässig dieser Partner ist, wie hoch die Fehlerquelle ist – weil wir jetzt das erste Mal außer Haus gehen. Bei Roboterjournalismus muss man ja auch bedenken, da gebe ich mich ja irgendwann auf ein dünnes Eis, wenn es automatisch publiziert wird. Das andere Produkt hatten wir im Haus entwickelt und hatten am Anfang auch wirklich vor dem Freischalten immer noch einmal einen Tipp eingebaut, dass wir die Meldung auch hätten löschen können. Nur, das würde dann nicht mehr gehen und deswegen muss es schon sehr zuverlässig sein. Natürlich kann immer mal ein technischer Fehler auftreten, aber es muss halt erst mal grundsätzlich ohne Fehler laufen. Das muss man erst mal abwarten, ob man das in der Form kann. #0:10:7.7#

KD: Genau. Thema Finanzen. Zu den ethischen Punkten komme ich noch. Denn gerade im Bereich der Finanzberichterstattung ist es viel, viel heikler, wer da die Haftung übernimmt, wenn Inhalte falsch sind. Was passiert, wenn jemand aufgrund dieser Aktienbeschreibung, die vielleicht aufgrund einer falschen Datenstruktur entstanden ist, einen Fehler macht? Dazu kommen wir noch. #0:10:33.2#

CE: Redakteure machen ja auch Fehler, oder? #0:10:34.9#

KD: Das wird das Argument sein, ja. #0:10:37.4#

CE: Ja, das ist die Frage: Wer macht weniger Fehler? Der Computer oder der Redakteur? #0:10:41.5#

KD: Aber dann ist ja auch die Frage der Manipulation: Wer kann die Datenbank manipulieren? Ein Mensch, der vielleicht für die Morgenpost arbeitet, ist vielleicht weniger manipulierbar (lachen) ... #0:10:55.2#

CE: ... Ja (lachen). #0:10:54.6#

KD: Also ich darf festhalten: Sie haben das erste Produkt inhouse entwickelt. Sie haben in der Testphase gesehen, dass es ganz gut läuft ... #0:11:1.6#

CE: ... auf jeden Fall haben wir gesehen, es ist Potential da. Es ist Potential da, weil es viele Bereiche im Journalismus gibt, die unglaublich arbeitsintensiv sind, die aber jetzt auch nicht so einen riesengroßen ... ja, weiß nicht ... was viele Kollegen auch nicht gerne machen. Muss man auch ehrlich sagen. Und wenn ich da bestimmte Sachen automatisieren kann und andere Kollegen können diese Arbeitskraft vielleicht in andere Dinge stecken, dann ist das eine tolle Innovation, die wir gut nutzen können. #0:11:44.5#

KD: Das ist der eine Punkt: Man kann Arbeit auf der einen Seite erleichtern. Auf der anderen Seite ist die Frage, was Sie in diesem Sinne damit bezwecken? Möchten Sie einfach die Schnelligkeit erhöhen, um Produkte und Inhalte zu haben und zu verteilen? Möchten Sie die Produktvielfalt erhöhen, um mit regionalen Inhalten präsent sein können, weil das natürlich im Datenjournalismus möglich ist? Es gab ja in der Kommunikationswissenschaft Studien, die schon Journalisten allgemein dazu befragt haben. Die sehen es alle entspannter als die Medien, die eher ängstlicher reagieren. Die sagen, es ist eine tolle Erleichterung, wenn mir jemand die faktenbasierten Finanznachrichten schreibt, weil es langweilig ist. Aber die Frage an Sie: Ist es wirklich nur fokussiert auf die Arbeitserleichterung oder geht es auch darum, dass Sie mit verschiedenen Produkten Monetarisierungsmöglichkeiten erkunden können, die vielleicht mit ihrer Personalstruktur nicht möglich wären? #0:12:46.1#

CE: Ich glaube, das wäre vermessen, darauf jetzt eine fertige Antwort zu geben. Wir sind ganz am Anfang einer Entwicklung bei der wir überhaupt noch nicht voraussehen können, wo die in drei oder fünf Jahren steht. Welches Potential hat überhaupt dieses Feld der automatischen Textgenerierung oder des sogenannten Roboterjournalismus? Wir sehen, dass es da interessante Role Models in den Staaten gibt. Wir sehen, dass einige große Nachrichtenagenturen damit arbeiten. Und die gefährden ja jetzt auch nicht vorsätzlich ihr Geschäftsmodell. Wenn Reuters oder The Associated Press damit arbeitet, dann kann man sich das schon mal ansehen. Und wir können davon auch bestimmte Elemente für unseren Markt und unsere journalistischen Arbeitsfelder adaptieren und das machen wir. Und wenn man das weiterdenkt: Welcher Journalist setzt sich denn drei oder fünf Jahre hin? Setzt sich der morgens hin und schreibt einen personalisierten Newsletter, indem der Leser oder Abonnent möglicherweise morgens gesagt bekommt: «Heute ist bei dir dieses Wetter. Das erwarten wir tagsüber bei dir und das ist an deinem Arbeitsplatz ... heute fährst du von dort nach dort zum Arbeitsplatz.» Das ist alles automatisiert in einem Newsletter, dann ist das schon eine unglaubliche Erleichterung und ein sehr großer Service für den User und für den Leser. Diese Investition lohnt sich. Man kann sehen was machbar ist. Das heißt ja nicht, dass da ein Roboter oder Algorithmus jetzt in der Zukunft eine Theaterkritik schreibt oder dass in Zukunft ein Kommentar über die Beschlüsse der EU-Außenminister durch einen Roboter oder Algorithmus geschrieben wird. Das ist Bullshit. Das

sind wirklich Schreckgespenster – vielleicht auch von interessierter Seite. Journalismus und Technik wachsen zusammen, das lässt sich nicht mehr trennen. Und man muss sehen, was dort möglich ist und wie man das verantwortungsvoll und innovativ in einem Markt einsetzen kann. Wir sehen das im Datenjournalismus, aber sicherlich auch im automatisierten Journalismus. Da gibt es viele Möglichkeiten, wo ich die Technik einsetzen kann. Aber wie gesagt, wir sind ganz am Anfang einer Geschichte. Wir müssen erst mal ausprobieren und dann können wir etwas dazu sagen: Brauchen wir das vielleicht nicht mehr oder können wir damit Geld verdienen? Wir testen im Moment ganz viel. #0:16:2.0#

KD: Ja, das ist ganz klar. Sie haben viele interessante Sachen gesagt, die ich auch auf dem Zettel habe. Kurz nochmal zusammengefasst über die Partnerschaft: Also Sie haben eine dreimonatige Testphase mit text-on zu einem Finanzprodukt. Haben Sie darüber hinaus schon mit anderen Anbietern gesprochen? #0:16:19.3#

CE: Nein. #0:16:19.9#

KD: Nur text-on? #0:16:21.4#

CE: Nach ihrem Input werde ich jetzt mit dem anderen über Sport reden (lachen). #0:16:33.4#

KD: Diese Inhouse-Entwicklung ... Wenn Sie sagen, Sie möchten innovationsgetrieben arbeiten: Können Sie entscheiden: «Ich möchte das ausprobieren?» oder wird das auf Managementebene gesteuert? #0:16:49.4#

CE: Ja. Wir entscheiden das hier. #0:16:53.6#

KD: Intern? #0:16:55.7#

CE: Ja. Intern. Das ist eine Entscheidung gewesen, die wir hier getroffen haben, und gesagt haben, wir investieren in diesen Bereich. Das ist eine Entscheidung, die wir vor zwei Jahren getroffen haben. Ich habe bewusst auf eine Planstelle oder zwei Planstellen an anderen Stellen verzichtet und habe gesagt, ich investiere hier in diesem Bereich. Ich möchte einen Web-Developer haben, ich möchte einen Datenjournalisten haben ... Ich habe jetzt einen zweiten Web-Developer dort sitzen. Und jetzt haben wir noch einen VJ, die hier zusammenarbeiten. Vielleicht habe ich so an einer anderen Stelle einen Journalisten weniger, aber da entsteht etwas Neues und das ist eine bewusste Strategieentscheidung gewesen. #0:17:36.8#

KD: Erstmal herzlichen Glückwunsch dazu. Wie würden Sie es einschätzen? Man spricht im Journalismus von diesem Schimpfwort «eierlegende Wollmilchsau». Der Journalist muss alles können. Inwieweit legen Sie den Fokus darauf, wenn man bei Ihnen ein Print-Volontariat machen will? Oder ist es vielleicht gekoppelt über die Funkegruppe Online-Print? Sind Programmierkenntnisse wichtig? Welchen Stellenwert haben sie bei Ihnen? #0:18:24.5#

CE: Ich glaube, wenn jemand heute Programmierkenntnisse hat, ist das so, als würde er früher ... Es ist ein weiteres sehr besonderes Talent, bei dem ich besonders zuhöre und dem ich besondere Aufmerksamkeit schenke. Wenn jemand Journalist werden will, ist es erstmal wichtig, dass er die Grund-Skills mitbringt. Er muss neugierig sein und ein gewisses Gefühl für die Sprache mitbringen. Und wenn mir jemand heutzutage sagt: «Ich bin in der digitalen Welt gar nicht unterwegs. Ich habe keinen Blog, bin nicht in den sozialen Netzwerken», dann bin ich skeptisch. Wenn mir daneben jemand sagt «Ich kann programmieren» und ich merke, ob er ein Sprachgefühl hat oder neugierig ist, und dann auch noch sagt «Ich kann programmieren», dann ist das ein gutes Match. So katapultiert der sich sicherlich einmal in die engere Auswahl. Es ist aber kein Must. Also nicht alle Journalisten sollten in Zukunft irgendwie programmieren. Ich glaube es ist wichtig, dass man darüber Bescheid weiß und dass man dieses Verständnis hat. Es gibt ja noch den Printbereich, der für viele Regionalverlage unglaublich wichtig ist: Unternehmen in diesem wachsenden Digitalstrang. Und da ist es glaube ich ganz wichtig. Was wir früher im Out an Grafikern gehabt haben, das sind jetzt andere Berufe. Das sind andere Skills, die da verlangt werden. Und das sind nun Entwickler. Leute die programmieren können. Genauso wie ich früher jemanden gehabt habe: «Mensch, ich möchte das Foto ein bisschen größer», brauche ich jetzt einen Programmierer der das, was ich mir ausdenke, wie ich die Geschichte erzählen möchte, umsetzt. Und ich weiß nicht, ob man da immer so eierlegende Wollmilchsäue anstellen kann oder ob das immer in einer Person ist. Ich glaube, es ist so, wie wir es hier haben. Wir haben den Glücksfall, dass wir einen Journalisten haben, der war eigentlich der Kern des Teams – mit dem hat das angefangen –, der programmieren kann und Journalist ist. Ein sehr guter Journalist mit Programmierkenntnissen. #0:20:44.8#

KD: Herr Tröger. #0:20:45.6#

CE: Julius Tröger. Mit dem fing das an. Dann hat er aber irgendwann gesagt «Wir machen schöne Geschichten, tolle Geschichten, können wir nicht mehr machen?». Dann sagte er «Ich bin eigentlich Journalist und ich kann programmieren, aber es gibt Leute, die können das viel schneller.». Dann haben wir uns entschieden zwei dazuzuholen, die eigentlich nur programmieren, aber journalistisches Interesse haben. Und man muss sagen, das war ein Treffer, ein Match, weil wir jetzt jemanden haben, der Sachen unglaublich schnell umsetzen kann. Das funktioniert nur als Team, wenn die beiden zusammenarbeiten. Es gibt eine führende Unternehmensberatung im Journalismus, die heißt *Innovation*. Sie sagen: «Wir werden in Zukunft verhältnismäßig drei zu eins arbeiten.». Drei Journalisten, ein Programmierer in den großen Newsrooms. Was mich wundert – und da sind wir ja ein kleines Haus und eine ganz kleine Marke – aber was mich wundert, dass die großen Marken das nicht konsequent umsetzen. Ich glaube, da gibt es im U.S.-Raum noch eine größere Bereitschaft. Es gibt in Europa immer noch das Gerangel, auch wenn ich mir das beim SPIEGEL angucke, ja ... Aber die Journalisten müssen sich immer die Kapazitäten erkämpfen und auf der

Produktseite holen, weil sie diese nicht frei zur Verfügung haben. Es gibt ein paar erste Ansätze des Umdenkens dazu. Aber das ist doch noch sehr durch die alten Verlagsstrukturen gekennzeichnet, wo die Produktseite immer an der Geschäftsführung einklinkt und die Journalisten auf der anderen Seite waren. dann hat man so ein bisschen um die Kapazitäten gerungen. Aber das ist eigentlich von gestern – nicht mehr zeitgemäß. #0:23:12.1# [Smalltalk] von #0:23:12.1# bis #0:23:14.7#

KD: Wie viele Leute haben Sie angestellt? Herrn Tröger, noch einen Programmierer ... Wer war noch an der Entstehung des Feinstaubmonitors beteiligt? #0:23:24.9#

CE: Wir haben zwei Programmierer und Julius Tröger. #0:23:28.3#

KD: Zwei Programmierer und Herr Tröger. #0:23:30.0#

CE: Bei uns steht auch immer drunter, wer dran beteiligt war. #0:23:33.3#

KE: Okay. Ja. Das habe ich gesehen. Da stehen noch ein paar weitere Namen darunter. #0:23:36.0#

CE: Kann sein. Wir haben immer noch einen Designer, der am Ende noch darüber guckt. David Wendler ... #0:23:43.9#

KD: ... Und die sind alle fest angestellt bei Ihnen? Vollzeit? #0:23:48.2#

CE: David Wendler ist bei uns in der Gruppe festangestellt. In dem Team arbeitet er. Der Moritz Klack ist ein fester Freiberufler. Julius Tröger ist festangestellt. Dann gehört noch der Max Boenke zu dem Team. Das ist der Videomann, den wir vor allem für die großen Datenanwendungen immer brauchen. Er ist fest angestellt. Nein, nein, also schon ein größerer Teil ist noch fest angestellt. #0:24:16.5#

KD: Und die Entscheidung für diese innovationsgetriebene Produktentwicklung ist vor zwei Jahren gefallen? Heißt das nur für die Berliner Morgenpost, nicht für andere Medien der Funke-Gruppe? Sondern ... #0:24:29.5#

CE: ... Da waren wir noch gar nicht bei Funke. Wir haben das begonnen ... #0:24:33.3#

KD: ... Bei Springer noch? #0:24:33.9#

CE: In der Endphase bei Springer, wo wir gesagt haben ... Wo ich mich mit Julius besprochen habe. Der Bereich Datenjournalismus wird so unterschätzt. Wir müssen da eigentlich investieren. Gerade im Lokalen hat es halt eine Chance, weil Datenjournalismus immer dann spannend wird, wenn er um mich herum passiert. Wenn ich mich irgendwo hinstellen kann mit dem Handy und sage «Hier bin ich, zeig mir den Rest.». Dann ist es spannend und dann hat es einen echten Zusatznutzen. Und da ist das entstanden. Jetzt muss man sagen, die letzten anderthalb Jahre haben wir dort auch eine gewisse Performance hingelegt, die auch ganz okay war. #0:25:13.6#

KD: Da würde ich Ihnen ganz zustimmen. Weil wir gerade beim Thema sind: Sie haben bei Springer begonnen, machen es jetzt hier aus ihren persönlichen Kontakten, die Sie so

haben. Haben Sie noch andere Blätter auf dem Schirm, die zu dem Thema Textgenerierung experimentieren und was machen und es antreiben? #0:25:32.3#

CE: Nein. Textgenerierung glaube ich ... Ich glaube Tomorrow Focus ist dort unterwegs. Ansonsten halten sich bei dem Thema alle bisher ... Also das wüsste ich nicht. Es kann aber sein, dass in letzter Zeit daran gearbeitet wird. Datenjournalismus schon, ja. Da gibt es zwei, drei in Deutschland, die mittlerweile auch sehr, sehr gut sind. Die ZEIT ist da sicherlich sehr gut. Der SPIEGEL fängt wohl so langsam an. Die NZZ ist sehr, sehr gut beim Datenjournalismus. Bei Robot ... weiß ich es nicht. Das würde ich jetzt am ehesten auch eher der NZZ zutrauen, weil die gerade im Datenbereich sehr gut sind und dort auch eigentlich ein sehr gutes Team haben. #0:26:21.1#

[Off the record) von #0:26:21.1# bis #0:27:31#

KD: In welchem finanziellen Rahmen bewegen wir uns mit der Investitionssumme in dem Text-on-Projekt? #0:27:31.7#

CE: Kleines Geld. #0:27:33.7#

KD: Kleines Geld heißt: 10.000 Euro, 20.000 Euro? #0:27:35.2#

CE: Wir bewegen uns in einem kleinen fünfstelligen Bereich. Nein, das wären für uns für drei Monate ... Wie gesagt, wir sind eine kleine Marke. Wir probieren aus. Ich habe jetzt auch nicht die unbegrenzten Kapazitäten und das ist ein Test. Wir sagen einfach mal, wir nehmen die börsennotierten Unternehmen und versuchen mal mit denen in Berlin zu arbeiten. Was können wir damit machen und was können wir auf der Plattform machen und läuft das verlässlich? Das kann auch sein, dass es überhaupt keinen interessiert. Dann schalten wir es auch wieder ab. Wir müssen auch erstmal sehen, ob es da ein User-Interesse gibt. #0:28:17.8#

KD: Aber live ist es noch nicht? #0:28:19.1#

CE: Nein. #0:28:21.5#

KD: Wird es in eine Live-Testphase gehen? Werden die Texte gekennzeichnet? #0:28:24.7#

CE: ... Ja, das steht dann immer da. #0:28:33.5#

KD: Das könnten Sie ja theoretisch wegnehmen, oder? Um zu sehen, wie oft auf einen Artikel geklickt wird. Wenn die laufen, ist das Interesse da. Das könnten Sie ja vergleichen ... z.B. haben Sie dann Klickauswertungen zum Feinstaubmonitor, dann könnten Sie sagen, das Projekt ist erfolgreich gewesen, je nachdem wie viele Leute auf die Feinstaubmonitor-Artikel geklickt haben ... #0:28:56.2#

CE: ... Man muss sagen, dass der Feinstaubmonitor jetzt nicht einer der tollsten interaktiven Anwendungen war, die wir jemals gemacht haben. Es war eher für uns ein Interessensprojekt. Kriegen wir das selbst hin? Kriegen wir die Textbausteine wirklich hin, dass das automatisch publiziert wird? Und das ging. Nun ist es so, dass das Thema Feinstaub nicht

jeden Tag jeden in Berlin interessiert. Und wenn jetzt irgendwie eine große Feinstaubbelastung da ist, dann gehen wahrscheinlich alle drauf. Aber ansonsten hält sich das eher in Grenzen und da haben wir andere Geschichten aus dem Team, die besser performen. Grundsätzlich würden wir immer dazuschreiben, dass es ein automatisch generierter Text war. Und ich würde es auch nicht aus Testgründen wegnehmen, dass es nicht automatisiert generiert ist. Denn im Netz zählen Transparenz und Glaubwürdigkeit. Man muss schon kenntlich machen, was die Quelle ist und wie das entstanden ist. Und man muss einem User da schon sehr klar und deutlich sagen, wie die Nachricht und wie dieser Text entstanden ist. Dann weiß er auch damit umzugehen. #0:30:22.5#

KD: Auf jeden Fall. Gut. Zu den anderen ethischen Herausforderungen komme ich noch. Allgemein zu den ökonomischen Faktoren, und zwar den Kosten für den Feinstaubmonitor. Gab es da außerhalb den allgemeinen personellen Kosten keine weiteren Kosten? Normales monatliches Gehalt? Wie lange hat der Entwicklungszeitraum gedauert, bis er ungefähr fertig war? #0:30:45.4#

CE: Drei, vier Monate. #0:30:49.7#

KD: Okay. Ein großer Punkt, der jetzt wahrscheinlich noch keine Rolle spielt, weil Sie in der Testphase sind, wären die Kosten für die Datenbanken. Also, Ihr Fokus liegt auf den frei zugänglichen APIs? #0:31:4.8#

CE: Nein, nicht unbedingt. Beim Thema Wirtschaft brauchen wir natürlich auch Daten. Aber diese Daten kaufen wir, weil wir dazu Daten für unsere Marke brauchen. Diese geben die Lieferanten weiter an text-on. Genau da reden wir mit unserem Lieferanten und sagen «Wir möchten die auch weiterverwerten.». Dann zahlen wir auch irgendwie fünf Euro oder einen kleinen Betrag mehr und können die Daten verwerten. #0:31:36.0#

KD: Wie heißt denn ihr Datenlieferant? #0:31:37.9#

CE: Ich glaube wir wechseln gerade von Bloomberg zu VWD. #0:31:46.9#

KD: Bloomberg macht ja auch etwas mit Automatisierung ... #0:31:50.4#

CE: ... Das weiß ich nicht. Wir waren bisher bei Bloomberg, weil wir über Axel Springer eingekauft haben, und jetzt wechseln wir zu VWD. #0:31:58.6#

KD: Und bei Axel Springer ist es, jetzt nachdem Sie weg sind, haben Sie ... Ich kann es mir fast gar nicht vorstellen. Ich habe die Presseleute bei Springer extrem genervt und Sie haben gesagt, sie hätten herumgefragt und es sei kein Thema. #0:32:11.1#

CE: Unglaublich. #0:32:13.5#

KD: Es ist ein Wahnsinn. Ich habe einen Kollegen, der ist von der Universität St. Gallen, der ist in der Springer-Journalistenschule und ist bei der BILD im Newsroom. Der hat keine Zeit jetzt, da etwas zu machen. Aber er sagt auch, dass es bei der BILD auf jeden Fall kein Thema ist. #0:32:31.4#

CE: Nur Reichweite. #0:32:31.8#

KD: Nur Klicks. #0:32:33.6#

CE: Richtig. Also es ist sehr reichweitengetrieben. Die haben sicherlich am ehesten die Möglichkeiten, das finanziell und personell auf die Beine zu stellen, wenn man die Ressourcen dort richtig einsetzen würde. #0:32:57.9#

KD: Der Feinstaubmonitor läuft auf Ihren hausinternen Servern? #0:33:3.3#

CE: Ja. #0:33:4.0#

KD: Der ist nicht ausgelagert? #0:33:5.0#

CE: Nein. #0:33:6.8#

KD: Wie wird es bei dem neuen Projekt sein? Wird das über den Server von text-on laufen oder wird es z.B. innerhalb Ihres CSM integriert? #0:33:14.8#

CE: Geplant ist eigentlich, dass es in unserem CMS integriert wird. #0:33:24.7#

KD: Im Augenblick schauen Sie einfach, ob die Texte, die text-on im Finanzbereich generiert den Qualitätsansprüchen der Berliner Morgenpost gerecht werden und wie variabel sie sind ... #0:33:37.6#

CE: ... Jetzt ist es eher, «wie kompatibel ist der Datenlieferant?». Also wie kompatibel ist ihre Technik mit unseren technischen Voraussetzungen. Dann testen wir es erstmal ohne etwas live zu schalten. Dann habe ich ja sicherlich auch Werte dabei ... Da kauft jemand irgendwie 30 Aktien und da ist sonst keine Bewegung und ich habe auf einmal einen riesen Ausreißer. Was mache ich denn damit? Dann muss ich mich irgendwie dazu verhalten. Oder ich muss es kenntlich machen oder ich muss bei dem Wert dazuschreiben, wie ich ihn behandle. Und das ist dann auch die Frage: Kann ich überhaupt aus den Werten einen Index ziehen? Weil, wenn ich dann so einen Wert dabei habe, der auf einmal 12 oder 13 Prozent hoch ist, reißt das dann auch mal den ganzen Index hoch. Da muss ich mich gerade mit diesen Fragen beschäftigen. #0:34:36.9#

KD: Okay, also können Sie noch keine Aussage darüber treffen. Das ist ebenfalls interessant. Man kann ja der Wissenschaft viel vorwerfen, beispielsweise, dass sie immer zu langsam ist. Aber aktuell sind wir eigentlich, gerade in den USA, auch die deutschsprachige Wissenschaft durch mich oder Kollegen aus München, hoch im Kurs. Die Münchner haben jetzt z.B. eine Studie durchgeführt zur Rezeption von automatisierten Texten. Die arbeiten mit Texten von Aexea, also in einer Kooperation mit Aexea. Sie haben festgestellt, dass User bei automatisierten Texten keine großen Unterschiede feststellen. Gerade bei der Kürze und im Sportbereich gibt es keine großen Unterschiede. #0:35:24.4#

CE: Wenn es gut gemacht ist, bin ich fest davon überzeugt. #0:35:26.4#

KD: Da gibt es auch die Studie von Clerwall, die das Ergebnis bestätigt. Zur Qualitätsfrage jetzt erst mal; es ist klar, dass das individuell angepasst werden muss, und da sind ja die Firmen ... #0:35:42.6#

CE: ... Wenn ich über kurze Nachrichtentexte rede, bin ich fest davon überzeugt. #0:35:47.8#

KD: Ich kann Ihnen gerne noch einen Artikel von mir geben, der einen Überblick schafft. Aber Natural Language Generation, was es in dem Sinne ist, das gibt es schon seit den fünfziger Jahren. Es ist halt jetzt erst für den Journalismus relevant geworden. Es gab bereits Berichte in den 1950er Jahren – das ist kein Hokuspokus. Aber wir sprechen erst dann von wirklichen Problemen, wenn eine Maschine bewerten muss. Aber es ist schwierig. Beispielsweise ist es im Reportagenbereich eher nicht möglich. Es braucht eher – auch was die Amerikaner sagen – eine Ausgangsbasis für Texte. Gerade dieser Quakebot... Ich habe mit Ken Schwenke gesprochen, einem Journalisten, der jetzt von der LA Times zur New York Times gewechselt ist. Er war der Erste, der sich damit beschäftigt hat. Deswegen heißt es in der Öffentlichkeit «Wow. Was macht er da?». Deswegen wird er so gehypt. In den USA gibt es ja Automated Insights und Narrative Science, die große Player sind. Wobei Narrative Science, durch die CIA-Geschichte sich da irgendwie vom Markt geschossen hat. Also in Deutschland sowieso. Das nur als Hintergrund. Ich habe – wie gesagt – einen Artikel geschrieben mit einem Marktüberblick. #0:37:58.6#

[Off the record: Exkurs The Guardian] von #0:37:58.6# bis #0:38:39.7#

KD: Meines Wissens sind Sie die einzige Zeitung, die sich da brandet. Sie branden sich auch als Innovationszeitung, weil die anderen Zeitungen den Trend verschlafen ... #0:38:51.3#

CE: ... Weiß ich gar nicht. Wir wollten es einfach ausprobieren. Das ist hier entstanden und dann haben wir das halt auch erzählt. #0:38:57.2#

[Smalltalk: Zeitmanagement, Stand des Interviews] von #0:38:57.2# bis #0:39:18.8#

KD: Wo, glauben Sie, können Produkte der Textgenerierung ihrem Leser der Berliner Morgenpost einen Mehrwert geben? Also, sind es die Bereiche, die Sie vorhin angesprochen haben, Servicebereiche, Sport und Wirtschaft? Oder wo sehen Sie den Mehrwert? #0:39:37.8#

CE: Ich glaube, dass die automatische Textgenerierung, wenn wir über personalisierte Nachrichten der Zukunft reden, eine entscheidende Rolle spielt. Und da gibt es halt unglaubliche Möglichkeiten. Ich habe das ja vorhin schon angeschnitten und ein Beispiel gesagt: Mein personalisierter Kosmos – den ich irgendwie zum Wissen brauche – der jeden Tag und jede Stunde, morgens, mittags, abends, aus Datenquellen neu generiert wird. Das ist schon ... Wenn ich das in einer intelligenten Form meinen Usern und Lesern anbiete, und dann noch auf Regionen beziehe, dann ist das ein Zusatznutzen, für den Leser auch

bezahlen. Das ist ein echt guter Service. Da kann ich dann vielleicht sogar Marketing einfließen lassen. Ich komme zwar nicht vom Marketing, aber da kann man ja Sales einbauen und sagen «Also du fährst von da nach da, also fährst du da lang, und wenn du da einkehrst, gibt es das und das.». Ich glaube in der Personalisierung, der automatischen Personalisierung von Services und Nachrichten, liegt echt eine große Chance. Es ist ein bisschen auch das, was die großen sozialen Netzanbieter anpeilen. Was macht Facebook? Facebook bietet das in gewisser Weise natürlich auch an. Aber natürlich bekommt man da auch nichts Genaues gesagt. Es sei denn, man hat die richtigen Leute abonniert oder hat sie in seiner Timeline. Aber oft bietet Facebook auch nicht die Daten an, die man braucht. Es ist ein interessanter Mix von Daten. Es ist zumindest ein Zusatznutzen, da muss man natürlich auch sehen, ob man den auch monetarisieren kann, ob der eine starke Loyalität erzeugen könnte. #0:41:56.1#

KD: Zur Kommunikation dieser Produkte: Hat die Berliner Morgenpost einen Betriebsrat? Was sagt der Betriebsrat dazu? #0:42:8.8#

CE: Super. #0:42:9.8#

KD: Findet er es gut? #0:42:11.9#

CE: Wir haben halt einen Betriebsrat, der sehr für Innovationen offen ist. Wir probieren halt viele Sachen aus. Deswegen fand ich auch irgendwie ... Es gab eine Zeile «Bei Funke sind jetzt die Roboter los.» Was für ein Bullshit. Wir probieren Dinge aus, und das ist doch nicht so, dass wir jetzt ein Computerprogramm ... Also das ist so als würde man mit wirklich fachfremden Augen auf eine Materie gucken von der man überhaupt keine Ahnung hat und dann auch noch darüber schreibt. Und das weiß der Betriebsrat auch. Es heißt doch nicht, wenn wir jetzt dort einen Feinstaubmonitor entwickeln, dass dort irgendwo Stellen wegfallen. #0:43:8.4#

KD: Haben Sie – nochmal zu den Klickraten vom Feinstaubmonitor – ungefähr eine Größenordnung? #0:43:13.7#

CE: Nein. Haben wir nicht. #0:43:15.2#

KD: Aber der läuft aktuell noch weiter und der wird weiterlaufen? #0:43:18.8#

CE: Ganz normal. #0:43:20.5#

KD: Und diese offene API ist von der Stadt Berlin irgendwie... teiloffen? Gehackt? (Lachen) #0:43:30.8#

[off the record bis #0:43:56.5#]

KD: Gut. Dann gehen wir noch mal die letzten zehn Minuten zu den ethischen Richtlinien: Tom Kent ist ein Managing Editor bei Associated Press, und beschäftigt sich auch mit dem Thema. Gerade hat er einen Guide zu den ethischen Richtlinien herausgegeben. Dort steht, wie Redaktionen verfahren sollten, wenn sie sich mit automatisierten Inhalten beschäftigen. «Tom Kent: Ethical Guidelines», kann ich Ihnen auch gerne noch schicken. Gibt es bei

der Berliner Morgenpost einen Code of Conduct – einen festgeschriebenen Code, wie Sie bei der Verwendung von Daten umzugehen haben? #0:44:45.6#

CE: Nein. Unser Code of Conduct und unser wichtigster Punkt ist, dass wir bei allen Datenanwendungen, ob sie automatisiert sind, ob wir Datentools online stellen, jedes Mal sehr transparent mitteilen, woher die Daten sind, wie wir sie verarbeitet haben, auf welcher Basis diese Daten erstellt worden sind. Wir haben in jeder Datenanwendung eine umfangreiche Legende – man kann auch ab und zu mal auf Fehler hinweisen. Man kann mit unseren Leuten kommunizieren. Wir rufen sogar dazu auf, uns Bugs und Fehler zu melden, um letztendlich besser zu werden. Wir gehen relativ offen damit um, wo wir etwas haben. Wir stellen auch teilweise danach unsere Datenanwendungen zur Verfügung. Wir haben schon Datenanwendungen und Tools entwickelt, die wir dann einfach auch zur Verfügung gestellt haben. Diese haben wir dann ins Netz gestellt oder bei GitHub veröffentlicht und gesagt «Arbeitet ihr auch damit.». #0:45:53.3#

KD: Tom Kent sagt, dass es wichtig ist, dass man sich auf die journalistischen Kriterien, die Objektivität der Daten weiter fokussiert. Was ja auch ganz klar ist. Er spricht da von der Kontrolle, von der Kontrollfunktion ... #0:46:8.6#

CE: ... Ich glaube, die gehört auch dazu, und deswegen werden zwei Webentwickler scheitern, wenn sie versuchen dort ein journalistisches Produkt auf den Markt zu geben. Beziehungsweise müsste man da dann sehr genau darauf achten, wie seriös das ist. Wichtig ist: Können wir diese Daten verwenden? Sind sie verwendbar, ist die Datenbasis richtig? Sind dort Fehler? #0:46:36.6#

KD: Wie können Sie das z.B. kontrollieren? #0:46:39.9#

CE: Dafür haben wir z.B. Julius Tröger, der im Bereich Statistik und Datenrecherche sich genau mit diesen Themen beschäftigt. Sind das eigentlich die Daten, die wir brauchen? Und stimmen die alle? #0:47:9.0#

[Smalltalk: Unterbrechung, Gliederung des Interviews] von #0:47:9.0# bis #0:47:14.3#

KD: Diese Kontrolle ist ganz interessant. Ich habe in einem Interview von Ihnen vom Vier-Augen-Prinzip gelesen. Die Associated Press hat jetzt die Kontrolle ihrer automatisierten Nachrichten eingestellt, weil sie sagen, sie schaffen es überhaupt nicht mehr, die zu kontrollieren, weil so schnell so viel publiziert wird. #0:47:33.4#

CE: Bei dem ersten Mal haben wir ein Vier-Augen-Prinzip verfolgt. Ich bin mir nicht sicher, ob wir das dort jedes Mal im Vier-Augen-Prinzip sicherstellen können, wenn jetzt das Börsentool irgendwann live gehen wird. Und das funktioniert auch so, dass man teilweise alle drei Minuten einen neuen Text hat. Das geht gar nicht mehr. Und da muss man einfach sehr klar darauf hinweisen, dass es automatisch generiert ist, und es darf keine riesen Bugs haben. Dafür gibt es aber auch einen Testlauf. Und, wie gesagt, auch Journalisten können Fehler machen. #0:48:15.5#

KD: Haben Sie sich da schon über die rechtlichen Konsequenzen Gedanken gemacht? Haftet text-on, wenn der Datensatz fehlerhaft ist? #0:48:24.0#

CE: Das ist ehrlich gesagt ein guter Punkt, den ich aus dem Gespräch mitnehme, dass ich das zumindest durch unsere Juristen überprüfen lasse. Wie weit wir dort in der Haftung sind oder wer, wann, wie uns dort in Regress nehmen kann aufgrund eines solchen Artikels ... Oder, ob wir vielleicht möglicherweise einen extra Vermerk darunter machen müssen. #0:48:44.3#

KD: Genau. Also, Sie haben noch keine Ergebnisse, was die Rechte an Daten, Rechte an den publizierten Texten aufgrund der Datenbasis... da haben Sie noch kein Dossier, wie es aussieht ... #0:49:3.3#

CE: Bei den lokalen Daten holen wir uns erst mal alles, was wir für sinnvoll halten und verwerten können. Auch wenn es schwer verfügbar ist. #0:49:15.5#

KD: Sie haben gesagt, nochmal zu der Kennzeichnung: Sie labeln alle und branden alle ihre automatisierten Texte. Ist es auch in Zukunft geplant, die Artikel, die automatisiert generiert werden, mit «Dieser Artikel wurde von text-on erstellt» zu beschriften? #0:49:28.7#

CE: Nein. #0:49:29.5#

KD: Oder «Dieser Artikel wurde automatisch ...». #0:49:31.8#

CE: ... Ja. Dann würde unten wieder die Legende stehen, wie das entstanden ist. Also das können Sie in jeder unserer Anwendungen sehen. Da würde man dann sagen «Zusammenarbeit mit text-on.». #0:49:45.0#

KD: Okay. Gut. Die Verantwortung ist auf einer rechtlichen Ebene, die noch geprüft wird. Aber Sie sagen, dass Sie in einen Dialog mit dem Leser treten und versuchen, die Produkte zu optimieren, oder? #0:50:4.5#

CE: Genau. Wir rufen jedes Mal auf, dass der Leser uns Feedback gibt, wenn wir ein neues Produkt oder ein Datentool launchen. Sie sagen uns, wo die Bugs sind, und wir versuchen dann sie abzuschalten. Das geht auch heutzutage gar nicht mehr anders. #0:50:18.1#

KD: Und diese Textgenerierung ist jetzt einfach nur ein weiteres Produkt im Trial-and-Error-Verfahren: Wie können praktisch den Journalismus für den Leser ... #0:50:27.2#

CE: ... Neue Monetarisierungsmöglichkeiten und wie können wir ... Was ist im Journalismus der Zukunft möglich? Genau. #0:50:39.4#

KD: Und da ist eben diese Textgenerierung in weitere Produkte integriert, die Sie zusätzlich noch im Digital Storytelling haben, Datenvisualisierungen usw. #0:50:51.7#

CE: Genau. Digital Storytelling ist bei uns auch ein großer Bereich. Aber das verändert sich ja auch. Die Zeiten, wo wir vielleicht 300, 500 oder 800 Zeilen Texte online gestellt haben, und dazwischen ein paar Fotos – das will ja heutzutage keiner mehr lesen. Oder wenn man

sich die Tiefe bei den großen Storytelling-Produkten, über die alle so toll berichtet haben, ansieht, wie viel da gelesen wurde. Das ist ja teilweise sehr ernüchternd. Und da versuchen wir natürlich auch immer wieder zu optimieren. Und wir versuchen vor allem den Aufwand, mit dem wir dann Digital Storytelling machen, so gering zu halten, dass er auch schnell umsetzbar ist. Man muss sagen können, okay das ist die Idee und dafür brauchen wir jetzt zwei Tage. Und nicht irgendwie ein Team von sechs Leuten, die dann irgendwie zwei, drei, vier Wochen oder Monate daran arbeiten. Also dann wären wir ja Lichtjahre von einer Monetarisierung entfernt. Wenn ich aber relativ einfache Tools habe, mit denen ich dann arbeiten kann, dann macht das auch Spaß. Davon haben wir uns jetzt zumindest mal einige Anwendungen selbst geschrieben und entwickelt. #0:52:10.7#

[Smalltalk: Abschluss] von #0:52:11# bis #0:53:9.0#

A.4 Tom Kent, Deputy Managing/Standards Editor, The Associated Press

Interviewer: Konstantin Dörr (KD)

Interviewpartner: Tom Kent (TK), Deputy Managing/Standards Editor, The Associated Press

Form des Interviews: Face to Face via Skype

Datum: 28.05.2015

Dauer: 00:28:42 Minuten

Sprache: Englisch

TK: Make sure that it exactly fits to what we want. #00:0:04#

KD: That is true. I read that the development process took over a year to fit into the AP's style guidelines. It is a long process. It's a product specifically developed for Associated Press. Right? #0:0:20.9#

TK: Right. #0:0:21.3#

KD: Automated Insights does not sell it to other clients in the media sector? #0:0:27.0#

TK: I do not believe so. I think we sell it to our subscribers. #0:0:35.0#

KD: Do you have any information on the costs of the programming – of the development of the product? #0:0:44.4#

TK: No. I am sorry again. I am an editorial guy. #0:0:50.2#

KD: Okay. About the code: Who is actually writing the code? The code for NLG was written by Automated Insights but in constant talks with the journalists of The Associated Press, am I right? #0:1:7.1#

TK: Right. We are talking about our editorial needs. We were not talking code. We were talking editorial needs and they would convert that into code. #0:1:17.7#

KD: Okay. So whom would you consider to be a journalist? Do you consider Automated Insights, who are the producers as a journalist? Or would you consider the Associated Press to be the producers of the texts? #0:1:42.0#

TK: Well ... That is a good question. But I would say the product is journalism and that is a collaboration. #0:1:53.2#

KD: Okay. So is it possible – in your opinion – that service providers can act as journalistic actors through their incorporation with journalistic media organisations and partners? #0:2:10.5#

TK: Well. I would not say that they necessarily know nothing to begin with. I mean they have ... You know, Automated Insights has been doing journalistic work, journalistic related work, producing journalistic products for some time now. Therefore, it is not as if they have no background in journalism. But the details of what the story looks like really

is up to us. They are messaging the data and using their artificial intelligence. But we are setting the journalistic standards. #0:2:55.3#

KD: But as it is a natural language generation, the final product has to be defined before generation. So the journalist has to have an idea about the product, about the final text, about the structure of the text before its generated and programmed by the programmers of Automated Insights, is it right? #0:3:15.8#

TK: Correct. We know what we want from a journalistic standpoint and then they arrange the program to treat the data to produce what we want. #0:3:25.8#

KD: Okay. Just a follow-up question to the one before: Is Automated Insights the only service provider and partner of the Associated Press in terms of natural language generation right now? #0:3:42.4#

TK: I believe so. As far as I know of ... Yes. #0:3:48.1#

KD: I think the whole problem about natural language generation of the data is... You have to have data sources, which are reliable and you can use for your natural language generation. Before we get to the ethical points about data sources ... Do you know if The Associated Press has any contracts with data brokers? Where does the Associated Press get its data? Do they buy the data? Or are there open APIs data? Open data? #0:4:29.5#

TK: In the case of these financial stories, we get the data from a company called Zacks. Zacks extracts this data from company reports and puts them into a structured form so that the algorithm can be used. #0:4:47.9#

KD: And for the sports application? #0:4:51.8#

TK: Well ... We will probably be getting them from a variety of sources. Probably leagues to a great degree. Or just some degree, I don't know. #0:5:6.7#

[Smalltalk: Background KD] von #0:5:6.7# bis #0:5:20.0#

KD: Do you know any other service providers besides Automated Insights, Narrative Science, Yseop and LinguaStat in the U.S. who has contracts with other media organisations? #0:5:38.6#

TK: In the U.S. only? #0:5:39.7#

KD: Or wherever you know something. #0:5:43.6#

TK: Well, there is Syllabs in France. #0:5:44.8#

KD: Yes, I talked to them already. #0:5:47.4#

TK: And they did that thing for Le Monde. Also, do you know *Ritzau*? The Danish news agency is doing some language processing for financial news. And I don't know who they're doing that with. #0:6:5.1#

KD: In Germany, there are a few companies. Finanzen100.de, it is a financial portal for stock market information. I know the Berliner Morgenpost ... #0:6:33.8#

TK: ... Oh and I think Sport-Informationen-Dienst which is part of AFP; I know they are doing some stuff. #0:6:41.2#

KD: Do you know any other organisations or agencies, which are using this technology for journalistic use cases? #0:7:22.6#

TK: Do you know Forbes? #0:7:23.6#

KD: Yes, I know them. #0:7:25.2#

TK: Okay. I think Forbes is using Narrative Science. And I do not know any others off hands. #0:7:32.5#

[Off the record] von #0:7:32.5# bis #0:8:09.2#

KD: The texts produced by Automated Insights and the Associated Press are primarily in English, right? Do you plan to produce in other languages as well? #0:8:19.8#

TK: Well, they are all only in English right now. We do operate in smaller languages, but I do not know of any plans yet to produce them in other languages. #0:8:30.7#

KD: And do you know something about the markets the Associated Press wants to focus on in the first place? English speaking markets? #0:8:48.7#

TK: I think we have been talking to some people in Canada. There is some possible use in Canada. You have to look at the right end of the telescope. Our goal is not to seek markets per se. Our goal is to reduce the amount of time that our journalists have to spend on boring repetitive work and to free our journalists to do more things that are interesting. Where we start looking is not: Which market looks good? We start with the question: Which journalist of ours is doing boring and repetitive things? It does make a difference, which market their doing it for. If we can release them, that is our goal. #0:9:43.8#

KD: Yes. That is a good thing. Today I talked to the Neue Zürcher Zeitung. I did a short presentation. I talked about the limitations and possibilities of natural language generation and where it is an advantage to use in journalism and what technology is available right now. Perhaps it is also interesting for you. Due to the limited time: Do you know anything about the price model? How is the Associated Press handling it? Is the Associated Press licensing or lending the stories to the subscribers or the clients? #0:10:34.3#

TK: No, I do not really know. Again, our goal here is not to – at least at this point – make a lot of new revenue but to free our journalists. #0:11:0.4#

[Smalltalk: Transition] von #0:11:0.4# bis 0:11:8.4#

KD: Talking about the ethical aspects of automated journalism. I am writing a paper about the ethical implications of robot journalism. Of course, I read your list: How media organisations have to shape the ethical standards. Is there a Code of Conduct available for the Associated Press, for the public? How are you handling robot journalism? #0:12:4.6#

TK: No. We have not drawn up something special. Most of the points are just extensions of what our policies have long been in terms of accuracy and transparency about our sources. That is why we have a by-line at the bottom of every story and a statement of Automated Insights. And it includes a link to an Automated Insights' page where we explain how our system works and so forth. It is really very much a question of implementing our existing principles in a new environment. #0:12:52.9#

KD: How do you cope with errors, which were produced by the software? I saw that they were published on the news platform as well. I did not see it in the checklist: How do you communicate errors of the generated software? #0:13:15.6#

TK: It is a good point. It really should be on the checklist. We correct them as if we correct every other story. We will run a new version – we call it a write-through – that says in a note at the top, that it corrects such and such. Moreover, we try to be very transparent about our corrections. If we do not discover the error until the next day. Then we will do a completely new story. We call it a corrective story. Saying the Associated Press incorrectly reported such and such and here is the correct version. I think that is another one of our general principles that we apply to journalism. #0:14:0.0#

KD: When we are talking about checking the information produced by the algorithms. I read that the Associated Press stopped checking the generated texts because there is just too many and it is taking too long for journalists to check every single text. How often and how regularly is a human journalist checking the texts? #0:14:29.2#

TK: I think we are checking and reviewing every day. The reason is we obviously want to check the work of the algorithms but the fact is, that everyday there are earning reports. There are many of these reports that we want to write more detailed versions of. The automation is giving us a quick 1, 2, 3, 4, 5-paragraph version. If it is an important company, our journalists want to do a new version of that story very quickly. In addition, in the course of that they are naturally going to be looking at the numbers to make sure that there are no mistakes in the original version. So we spot check some just to make sure the machine is working right. It is not like that these exist in some world apart of our journalists. They are very much involved in this coverage all the time. And they are often building and advancing the stories that were created automatically. #0:15:40.8#

KD: Of course. So you would say that texts are just a starting point for further writing, visualization and quoting ... #0:15:52.6#

TK: Well, many of the texts or the companies are not that big or important, that we have to do anything with them. Nevertheless, if we would have the human resources to write longer versions of everything ... But texts on important companies: Yes, they are just the starting point. #0:16:10.6#

KD: Just what comes to my mind: How many clients are actually subscribing the service for the earning reports for example? Do you have a rough number? #0:16:25.3#

TK: I do not. I am not even sure that we sell these earning reports as a separate service. They just may be part of a service. At The Associated Press, it is just an additional thing we put in there. Because when we are doing them manually, we did not charge extra. It was just a part of The Associated Press' financial news service. I suspect these are just a part of the news service as well. I mean you can always imagine new figurations but I think that the goal is to serve subscribers better and take pressure off our journalists. #0:17:10.4#

KD: Just one last question about number four of your ethical checklist. You are writing or asking the question: «Does the style of the automated reports match your style?» The style of the media organisation? That was not quite clear to me. What did you mean with style? Do you mean the writing style? Or do you mean – what is not on the checklist – the principles, the rules and norms of how journalists should act in terms of objectivity, in terms of research, meaning for example the principles of two sources, of three sources and so on? #0:17:55.8#

TK: Well, of course these stories are one-source stories. They just report what the company reported. But yes, it includes all of that. I mean certainly from an objectivity standpoint we want these stories to be objective, we want them to be consistent in terms of how to treat data from one company and from another company. We want them to work by rules that are not skewed in depending on what company it is. There is some question, which literally meant by style: the punctuation, capitalization styles and so on. Then there is just the writing style. For example, you might write, «a company's revenue increased five times» or you might say «a company's revue shot up five times» or «a company's revenue shot investors by...». You know, there is also a degree of style in that sense. And our style generally is not to be sensationalist. We are going to face that a lot with sports. Because there's a tendency to do a lot of sensational sports writing. And we don't do it. And a lot of people do it. So when we're dealing with Automated sports texts we're going to have to make sure that it's written in a careful journalistic style. #0:19:41.2#

KD: And the sports application is developed by Automated Insights as well? #0:19:46.1#

TK: That is what we are working with right now, yes. #0:19:52.6#

KD: Another short question about the copyright, because I talked to Claude de Loupy from Data2Content from Syllabs. He said that depending on the copyright laws of each country, you have to face different issues. He told me that in France the author of a journalistic text is the copyright-holder and not the media organisation. So robot journalism has to face issues of copyright and who is the actual producer of the text. I am thinking about financial information – especially in Germany: How to get along if somebody, like an investor, relies his information about his latest stock market purchases on information that was produced

based on wrong data, from a service provider like Automated Insights? Who is to blame?
#0:21:6.0#

TK: Well, that is not copyright. That is a separate question. I mean a defamation or incorrect material is one issue – copyright is another. If it is a question of ownership of copyright... If you look in the legal literature, you can see arguments saying it is the news organisations. You can see arguments saying it is the person who provided the data. And then you claim it is the person who did the algorithm. I think the only thing you can say for sure is that – in our opinion – it is copyrighted to begin with by somebody. So people should not be going about taking it for free. It is copyrighted and it is protected. Then in terms of who exactly is the copyright-holder it is going to vary from situation to situation depending on the agreements that are being made between the data providers and the algorithm people and the news organisation. I think in their contracts, it is likely that they will – if there is any doubt – assign the copyright to one or another party. I think that they will determine that. #0:22:30.8#

[Time management] #0:22:30.8# bis #0:23:6.8#

KD: Who do you think is the best person I could talk to about the whole business development of the Associated Press? Should I ask the communication ... #0:23:19.1#

TK: ...Yes! You should ask corporate communications department. #0:23:22.2#

[Off topic talk]

KD: How are your personal relations to robot journalism? Does it affect your daily work?
#0:24:5.9#

TK: Well, from day to day, it is not usually an issue. Usually I get involved when we are considering a new robot journalism project or when we are reviewing on how it is working. But I'm not involved every day. #0:24:27.9#

KD: Where do you personally see the possibilities and limitations? You talked about the possibilities of how the software is able to free journalists from boring work. Where can you see, from your experience, the limitations of robot journalism? #0:24:49.9#

TK: To talk about now. I cannot talk about where we will be in 20 years. But at the moment it is best used for relatively simple assignments. But simple work can make a dramatic difference for our staff and our readers. In terms of our staff, we already talked about what our goal is there, if we go to the readers, it is giving them coverage of things, which we would never be able to provide a coverage of. We are just not ... We would never be in a position to do these many earning reports as we do. We would never be in a position to cover all these sports leagues and so on. #0:25:46.1#

KD: So it is an additional journalistic product? #0:25:49.7#

TK: Yes. It is an additional journalistic product we would not be able to provide otherwise. Fortunately, we have the expertise. We have the editorial judgment and the editorial experience and this enables us to make this material really good. As good as it can possibly be. Our professional skills are a tremendous asset here. But this technology enables us to really scale up our journalistic capabilities and sensitivities to new areas which we are just beginning to discover. #0:26:44.2#

KD: So many service providers have issues in generating revenue. That is a problem. Many providers I spoke to told me that journalism is just one part of their business model. They will not rely on just producing journalistic products, because the market is in too much of a crisis right now. #0:27:43.4#

TK: Right. Well, I think Narrative Science and Automated Insights are also doing a lot of non-journalistic stuffs or at least they have an eye on it. This is just one branch that excites us and excites you but I do not think it is necessarily the core of their work. In fact, we do a lot of work for them and help them to train these arguments. #0:28:8.5#

[Smalltalk] #00:28:08# bis #0:28:42.0#

B. Ausweis der Eigenleistung

Artikel III – When Reporters get Hands-On with Robo-Writing. Professionals Consider Automated Journalism’s Capabilities and Consequences (Thurman, N., Dörr, K., & Kunert, J.)

Die Übersicht zeigt jeweils den Schwerpunkt der einzelnen Autorenleistungen innerhalb des Projekts. Feedback zu spezifischen Problemen wurde – je nach Anforderung – von allen drei Autoren auf jeder Projektstufe gegeben.

Tabelle 2: Überblick Autorenleistung, Artikel III

Artikel III	Autor 1 (Thurman, N.)	Autor 2 (Dörr, K.)	Autor 3 (Kunert, J.)
Projektidee & Entwicklung der Fragestellung	X	X	x
Projektorganisation & Kommunikation	X	X	x
Projektfinanzierung	X	x	
Literaturrecherche	x	X	x
Entwicklung des Untersuchungsdesigns & Methode	X	X	X
Datenerhebung (Workshop: Beobachtung & Interviews)	X	X	X
Datenanalyse & Interpretation	X	X	
Konzeption & Schreiben des Manuskripts	X	x	
Einreichung & Überarbeitung des Manuskripts	X	x	

Artikel IV – Normative Qualitätsansprüche an algorithmischen Journalismus (Dörr, K., Köberer, N., & Haim, M.)

Die Übersicht zeigt jeweils den Schwerpunkt der einzelnen Autorenleistungen innerhalb des Projekts. Feedback zu spezifischen Problemen wurde – je nach Anforderung – von allen drei Autoren auf jeder Projektstufe gegeben.

Tabelle 3: Überblick Autorenleistung, Artikel IV

Artikel IV	Autor 1 (Dörr, K.)	Autor 2 (Köberer, N.)	Autor 3 (Haim, M.)
Projektidee & Entwicklung der Fragestellung	X	x	X
Projektorganisation & Kommunikation	X		
Literaturrecherche	X	x	X
Konzeption & Schreiben des Manuskripts	X	x	
Einreichung & Überarbeitung des Manuskripts	X		

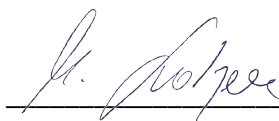
Artikel V – Ethical Challenges of Algorithmic Journalism (Dörr, K., & Hollnbuchner, K.)

Die Übersicht zeigt jeweils den Schwerpunkt der einzelnen Autorenleistungen innerhalb des Projekts.

Tabelle 4: Überblick Autorenleistung, Artikel V

Artikel V	Autor 1 (Dörr, K.)	Autor 2 (Hollnbuchner, K.)
Projektidee & Entwicklung der Fragestellung	X	X
Projektorganisation & Kommunikation	X	
Literaturrecherche	X	X
Konzeption & Schreiben des Manuskripts	X	x
Einreichung & Überarbeitung des Manuskripts	X	

Hiermit bestätige ich die Eigenleistung des Autors bei Gemeinschaftspublikationen gemäß §7 Abs. 3 der Promotionsverordnung der Philosophischen Fakultät der Universität Zürich vom 8. Juli 2009.



Prof. Dr. Michael Latzer
Hauptverantwortliche Betreuungsperson

Zürich, 15.05.2017

C. Lehr- und Forschungstätigkeiten

Die folgenden Abschnitte geben einen chronologischen Überblick über die Lehrtätigkeiten an der Universität Zürich (2013–2017), Präsentationen auf Konferenzen, Workshops und der Teilnahme an Panels (2010–2017).

C.1 Lehre

HS 13

- SP1-Forschungsseminar: Gesellschaftliche Implikationen von Suchmaschinen, Newsaggregatoren & Co. (zusammen mit Katharina Hollnbuchner) (Teil 1)

FS 14

- SP1-Seminar: Auswirkungen von Medieninnovationen: Privatheit am Ende?
- SP1-Forschungsseminar: Gesellschaftliche Implikationen von Suchmaschinen, Newsaggregatoren & Co. (zusammen mit Katharina Hollnbuchner) (Teil 2)
- VL: Facetten des Medienwandels (zusammen mit Michael Latzer, Joscha Gewinner, Katharina Hollnbuchner, Natascha Katharina Just, Julian Wallace)

HS 14

- SP1-Forschungsseminar: Gesellschaftliche Implikationen von Suchmaschinen, Newsaggregatoren & Co. (Teil 1)

FS 15

- SP1-Forschungsseminar: Gesellschaftliche Implikationen von Suchmaschinen, Newsaggregatoren & Co. (Teil 1)
- SP1-Seminar: Big Data, Big Brother, Big Business – Privacy im Wandel?

HS 15

- Grundlagen der Publizistik- und Kommunikationswissenschaft I (zusammen mit Michael Latzer und Natascha Just)
- Grundlagen der Publizistik- und Kommunikationswissenschaft II (zusammen mit Michael Latzer und Natascha Just)

FS 16

- SP1-Seminar: Big Data, Big Brother, Big Business – Privatheit im Wandel?

HS 16

- SP1-Forschungsseminar: Gesellschaftliche Implikationen von Suchmaschinen, Newsaggregatoren & Co. (Teil 1)

FS 17

- SP1-Forschungsseminar: Gesellschaftliche Implikationen von Suchmaschinen, Newsaggregatoren, & Co. (Teil 2)
- VL: Facetten des Medienwandels (zusammen mit Michael Latzer, Moritz Büchi, Benjamin Gerwoll-Ronca, Julian Wallace)

Insgesamt betreute Bachelor-Arbeiten: 39

Insgesamt betreute Master-Arbeiten: 2

C.2 Präsentationen, Panel und Workshops

- Thurman, N., Dörr, K., & Kunert, J. 2017. *When Reporters get Hands-On with Robo-Writing: Professionals Consider Automated Journalism's Capabilities and Consequences*. Vortrag auf der International Communication Association's (ICA) 67th Annual Conference, San Diego, CA.
- Thurman, N., & Dörr, K. 2016. *When Reporters Get Hands-On with Robo-Writing*. Vortrag auf der Veranstaltung «News – Automaten. Journalismus in Zeiten von Big Data, Algorithmen und künstlicher Intelligenz» des Internet & Gesellschaft Collaboratory, Google Deutschland, Berlin, Deutschland, 6. Dezember 2016.
- Dörr, K. 2016. Panelteilnehmer auf der Veranstaltung «News – Automaten. Journalismus in Zeiten von Big Data, Algorithmen und künstlicher Intelligenz» des Internet & Gesellschaft Collaboratory, Google Deutschland, Berlin, Deutschland, 6. Dezember 2016.
- Dörr, K. 2016. *Algorithmen: Die Journalisten der Zukunft?* Vortrag im Diskussionsforum: Künstliche Intelligenz für Arbeit, Gesundheit und Alltag – Zukunft mit unbegrenzten Möglichkeiten? Collegium Helveticum, ETH Zürich.
- Haim, M., & Dörr, K. 2016. *Automatisierte Qualität? Zur Institutionalisierung und normativen Qualität von algorithmischem Journalismus*. Vortrag auf der Jahrestagung der Deutschen Gesellschaft für Publizistik und Kommunikationswissenschaft (DGPuK). Leipzig, Deutschland, 30. März bis 1. April 2016.
- Dörr, K. 2015. *Algorithmic Journalism. Insights, Challenges and Prospects of an Automated News Production for Journalism Research*. Vortrag auf dem Symposium Change in Journalism, Schloss Herrenhausen, Hannover, Deutschland, 16. bis 18. September 2015.
- Dörr, K., & Hollnbuchner, K. 2015. *Ethical Challenges of Algorithmic Journalism*. Vortrag auf The Future of Journalism Conference: Risks, Threats and Opportunities. Cardiff University, GB, 10. bis 11. September 2015.
- Dörr, K. 2015. *Algorithmic Journalism. Theoretical, Economic and Technological Implications of Automated News Production in Journalism*. Vortrag auf der Shaping the Future of News Media Conference, International Conference on Integrated Journalism Education, Research and Innovation. Barcelona, Spanien, 17. bis 19. Juni 2015.
- Dörr, K. 2015. *Algorithmic Journalism, Concept and Consequences*. Vortrag auf der Re-Inventing Journalism Conference (ECREA & DGPuK). Winterthur, Schweiz, 5. bis 6. Februar 2015.

- Godulla, A., Wolf, C., Hohlfeld, R., & Dörr, K. 2015. «We can all make great things» *The Potential of Crowdfunding for the Re-Invention of Journalism*. Vortrag auf der Re-Inventing Journalism Conference (ECREA & DGPUK). Winterthur, Schweiz, 5. bis 6. Februar 2015.
- Dörr, K. 2015. *The Future of Journalism*. Workshop auf der clunc: Computational Linguistics Unconference. Stuttgart, Deutschland, 23. bis 24. Januar 2015.
- Hollnbuchner, K., & Dörr, K. 2014. *Ethical Challenges of Algorithmic Selection*. Vortrag im Panel: Communication shaped by Algorithms: technological, economic, social and ethical Challenges an der Jahrestagung der European Communication Research and Education Association (ECREA). Lissabon, Portugal, 12. bis 15. November 2014.
- Wallace, J., & Dörr, K. 2014. Beyond Traditional Gatekeeping. How Algorithms and Users Restructure the Online Gatekeeping Process. Vortrag auf der Digital Disruption to Journalism and Mass Communication Theory Conference. Brüssel, Belgien, 2. bis 3. Oktober 2014.
- Dörr, K., & Hohlfeld, R. 2013. *Zukunft der Kommunikationsberufe*. Vortrag auf den Kommunikationswissenschaftlichen Tagen – KWT 2013 – der Österreichischen Gesellschaft für Kommunikationswissenschaft (ÖGK): «Changing Communications II: Herausforderungen des gesellschaftlichen, medialen und kulturellen Wandels». Klagenfurt, Österreich, 17. bis 19. Oktober 2013.
- Dörr, K. 2010. *Teilnahme an der interdisziplinären Tagung «Privatheit»*. Universität Passau, Passau, Deutschland, 19. bis 20. November 2010.

C.3 Weitere Publikationen während des Doktorats

- Dörr, K. 2018. Journalist vs. Algorithm? Ethical Approaches to Computational Journalism. In *The Routledge Handbook of Developments in Digital Journalism Studies*, herausgegeben von Eldridge, S., & Franklin, B. [forthcoming].
- Caswell, D., & Dörr, K. 2017. Automated Journalism 2.0: Event-Driven Narratives. From simple Descriptions to real Stories. *Journalism Practice* [online first]. doi: [10.1080/17512786.2017.1320773](https://doi.org/10.1080/17512786.2017.1320773)
- Dörr, K. 2016. Algorithmic Journalism (Algorithmischer Journalismus). In *Journalistische Genres*, herausgegeben vom Deutschen Fachjournalisten-Verband (DFJV), 241–252. Konstanz: UVK.
- Latzer, M., Saurwein, F., Dörr, K., Just, N., Wallace, J. 2015. Jugendmedienschutz: Selbstregulierung der Schweizer Branchen auf dem Prüfstand. *Soziale Sicherheit, CHSS*, 4, 220–223.
- Latzer, M., Saurwein, F., Dörr, K., Just, N., Wallace, J. 2015. Protection de la jeunesse face aux médias: analyse de l'autorégulation en Suisse. *Sécurité sociale, CHSS*, 4, 220–223.
- Latzer, M., Saurwein, F., Dörr, K., Just, N., Wallace, J. 2015. *Evaluation der Selbstregulierungsmaßnahmen zum Jugendmedienschutz der Branchen Film, Computerspiele, Telekommunikation und Internet*. Forschungsbericht, 11/15, im Auftrag des Bundesamtes für Sozialversicherungen (BSV). Universität Zürich, Zürich.

D. Curriculum Vitae

Konstantin Nicholas Dörr

Universität Zürich
IPMZ – Institut für Publizistikwissenschaft und
Medienforschung
Andreasstrasse 15
CH-8050 Zürich

Büro: AND 3.42
Tel.: +41 (0)44 635 20 35
E-Mail: k.doerr@ipmz.uzh.ch

Persönliche Daten

Geburtsdatum: 7. November 1985
Geburtsort: München, Deutschland

Ausbildung

Seit 03/2013	Wissenschaftlicher Assistent am Institut für Publizistikwissenschaft und Medienforschung, Abteilung «Medienwandel & Innovation» (Prof. Dr. Michael Latzer), Universität Zürich Dissertationsprojekt: «Algorithmischer Journalismus. Eine Analyse der automatisierten Textproduktion im Journalismus auf gesellschaftlicher, organisatorischer und professioneller Ebene»
10/2010 bis 02/2013	Studium Medien und Kommunikation (M.A.), Universität Passau Thema der M.A.-Arbeit: «Zwischen Nachrichtenbürokratie und Boulevard. Eine Untersuchung zur Nachrichtenselektion in den Hauptnachrichten des deutschen Fernsehens»
04/2008 bis 07/2010	Studium Medien und Kommunikation (B.A.), Universität Passau Thema der B.A.-Arbeit: «Kommunikationsprognostik in der Mobilkommunikation. Eine quantitative Online-Befragung zur Zukunft, Chancen und Trends von Applikationen für Smartphones»
10/2005 bis 10/2007	Studium der Rechtswissenschaften, LMU München
09/1996 bis 07/2005	Herder-Gymnasium Forchheim Abschluss: allgemeine Hochschulreife

Auszeichnungen

10/2011 bis 10/2012	Deutschlandstipendium
07/2010	B.A.-Studium mit Auszeichnung



Erklärung

Hiermit erkläre ich, dass die Dissertation von mir selbst ohne unerlaubte Beihilfe verfasst worden ist und diese Dissertation noch an keiner anderen Fakultät eingereicht wurde.

Ort und Datum

Unterschrift

Zürich, 15.05.2017

Konstanze Dörr
